

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 19, 2004, 15:50:28 ; Search time 97.9924 Seconds

(without alignments)
2185.588 Million cell updates/sec

Title: US-09-903-199-2

Perfect score: 4022
Sequence: 1 MAQRKAKSSGSSGSGSGS.....IVDVHPELTPQQRSLUPAI 758

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

1: Geneseq_29Jan04:*
2: geneseqp1980s:*
3: geneseqp1990s:*
4: geneseqp2000s:*
5: geneseqp2001s:*
6: geneseqp2002s:*
7: geneseqp2003as:*
8: geneseqp2003bs:*
9: geneseqp2004s:*

Fred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4022	100.0	758	4 AAB83919	Aab83919 A human a
2	4022	100.0	758	6 ABG72365	Abg72365 Human asp
3	4022	100.0	758	6 ADA00639	Ada00639 Human asp
4	1353.5	33.7	255	6 ABU92053	Abu92053 Human pro
5	1334	33.2	255	6 AAY33642	Aay33642 Human lab
6	1320	32.8	255	5 AAU85544	Aau85544 Clone #48
7	1320	32.8	255	6 ABU69516	Abu69516 Human lun
8	1320	32.8	255	6 ABU66419	Abu66419 lung canc
9	933	23.2	422	4 AAB61986	Aab61986 Drosophila
10	274.5	6.8	359	4 AAB73682	Aab73682 Human oxi
11	271.5	6.8	359	4 AAB73682	Aab73682 Human oxi
12	262	6.5	226	6 AAB68322	Abm68322 Phototrab
13	241	6.0	104	4 AAU29679	Aau29679 Novel hum
14	218.5	5.4	299	6 ABU21099	Abu21099 Protein e
15	217.5	5.4	324	4 AAU80861	Aau80861 Novel hum
16	215	5.3	47	4 AAU31979	Aau31979 Novel hum
17	204.5	5.1	311	4 ABB60327	Abb60327 Drosophila
18	203.5	5.1	562	2 AAR70491	Aar70491 Leucocyto
19	199	4.9	109	4 AAB73672	Aab73672 Human oxi
20	199	4.9	783	2 AAR05804	Aar05804 C-termina
21	197.5	4.9	1018	2 AAR98747	Aar98747 P. vivax
22	197.5	4.9	1018	2 AAW97039	Aaw97039 A. secreta
23	197.5	4.9	1018	4 AAG65528	Aag65528 Plasmidin
24	197.5	4.9	1132	2 AAR37866	Aar37866 Chicken I
25	196.5	4.9	355	6 ABU23196	Abu23196 Protein e

26	191.5	4.8	1881	3 AAY44506	Aay44506 Streptoco
27	191.5	4.8	1881	6 ABU10447	Abu10447 S. pneumo
28	190.5	4.7	778	3 AAG46504	Aag46504 Arabidops
29	190.5	4.7	788	4 ABB68264	Abb68264 Drosophila
30	190.5	4.7	1162	3 AAY96255	Aay96255 Kaposi's
31	190.5	4.7	1162	3 AAY58500	Aay58500 HHV8 ORF
32	190.5	4.7	1162	4 AAB62331	Aab62331 Amino aci
33	190.5	4.7	1162	5 ABB05621	Abb05621 Kaposi's
34	190	4.7	380	2 AAM03627	Aam03627 Human fol
35	189.5	4.7	304	6 ABU17264	Abu17264 Protein e
36	189.5	4.7	312	6 ADA36728	Ada36728 Acinetoba
37	189.5	4.7	1616	6 ABU35669	Abu35669 Protein e
38	189.5	4.7	1616	7 ABO23515	Abu23515 Mycoplasma
39	189.5	4.7	3263	4 ABB67210	Abb67210 Drosophila
40	187.5	4.7	1192	4 ABB59642	Abb59642 Drosophila
41	186.5	4.6	412	2 AAW03626	Aaw03626 Human thy
42	186	4.6	49	6 ADA40577	Ada40577 Human sec
43	186	4.6	49	6 ABR47638	Abu47638 Human sec
44	186	4.6	49	6 ABR47638	Abu47638 Human sec
45	186	4.6	50	2 AAY01148	Aay01148 Secreted

ALIGNMENTS

RESULT 1	AAB83919	standard; protein; 758 AA.
ID	AAB83919	
XX	AAB83919;	
XX	23-JUL-2001 (first entry)	
DT	23-JUL-2001 (first entry)	
XX	A human aspartyl [asparaginyl] beta-hydroxylase (HAAH).	
DE	A human aspartyl [asparaginyl] beta-hydroxylase (HAAH).	
XX	Epidermal growth factor-like domain; EGF-like domain; cancer;	
KW	human aspartyl beta-hydroxylase; HAAH; malignant neoplasm; tumour.	
XX		
OS	Homo sapiens.	
XX		
PN	WO200135102-A2.	
XX		
PD	17-MAY-2001.	
XX		
PF	08-NOV-2000; 2000WC-US030738.	
XX		
PR	08-NOV-1999; 99US-00436184.	
XX		
PA	(RHOD-) RHODE ISLAND HOSPITAL LIFESPAN PARTNER.	
XX		
PI	Wands JR, De la Monte SM, Ince N, Carlson RI;	
XX		
DR	WFI; 2001-329171/34.	
XX		
DR	N-PSDB; AAF89811.	
XX		
PT	diagnosing malignant neoplasm in a mammal, involves contacting mammalian	
PT	sample with antibody that binds to human aspartyl beta-hydroxylase	
PT	polypeptide to form antigen-antibody complex and detecting the complex.	
XX		
XX	Disclosure; Page 5; 76pp; English.	
XX		
CC	The present sequence represents a human aspartyl (asparaginyl) beta-	
CC	hydroxylase (HAAH) enzyme. Epidermal growth factor (EGF)-like domains of	
CC	polypeptides are hydroxylated by HAAH enzymes. HAAH is used in the method	
CC	of the invention. The specification describes a method for diagnosing a	
CC	malignant neoplasm in a mammal. The method comprises contacting a body	
CC	fluid with an antibody which binds to HAAH polypeptide under complex	
CC	forming conditions, and detecting the antigen-antibody complex. The	
CC	method is useful for diagnosing and prognosing a malignant neoplasm in a	
CC	body fluid e.g. central nervous system (CNS)-derived body fluid, blood,	
CC	serum, urine, saliva, sputum, lung effusion, and ascites fluid of mammal,	
CC	where the neoplasm is derived from endodermal tissue and is selected from	
CC	colon cancer, breast cancer, pancreatic cancer, liver cancer, cancer of	

CC bile ducts, and cancer of the CNS, HAAH antibodies, linked to a cytotoxic
 CC agent, are useful for killing tumour cells
 XX
 XX Sequence 758 AA:

Query Match 100.0%; Score 4022; DB 4; Length 758;
 Best Local Similarity 100.0%; Pred. No. 8,1e-311;
 Matches 758; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 MAQRNNAKSSGNSGSSGSGSTASGSSPGARRRTKKGKGNKRGKGLSGTSFTPTMV 60
DB 1 MAQRNNAKSSGNSGSSGSGSTASGSSPGARRRTKKGKGNKRGKGLSGTSFTPTMV 60
QY 61 IALGVMTSVAVMWEDLVYEEVGLKGLIYDADGGDVPDADAKLGLKRSSTSEPAVP 120
DB 61 IALGVMTSVAVMWEDLVYEEVGLKGLIYDADGGDVPDADAKLGLKRSSTSEPAVP 120
QY 121 PEEAEPHTPEEQVVEAEPONIEDAKEQIQLSLHEWHAHEVGEEDLQOEDGPTGPQ 180
DB 121 PEEAEPHTPEEQVVEAEPONIEDAKEQIQLSLHEWHAHEVGEEDLQOEDGPTGPQ 180
QY 181 QEDDEFLMATVDVDRFETLEPEVSHETESHVHEETVSQDQNDMEBMSQENPDSE 240
DB 181 QEDDEFLMATVDVDRFETLEPEVSHETESHVHEETVSQDQNDMEBMSQENPDSE 240
QY 241 PVDEBERLHDDVDVTVVVEEQAVYEBLENGEITETVAPREDNPVEDSOVIVEEYSI 300
DB 241 PVDEBERLHDDVDVTVVVEEQAVYEBLENGEITETVAPREDNPVEDSOVIVEEYSI 300
QY 301 FVVEBQEQVPPETNNKTDPEQOKAVKVKKKKPKLANKPKDTIKAEIDAAEKLRKGLIEA 360
DB 301 FVVEBQEQVPPETNNKTDPEQOKAVKVKKKKPKLANKPKDTIKAEIDAAEKLRKGLIEA 360
QY 361 VNAPELVYKTPQSPRARVGAQCEDDIAEKRSNEVLRGALIEYQVAVSLPVPADILK 420
DB 361 VNAPELVYKTPQSPRARVGAQCEDDIAEKRSNEVLRGALIEYQVAVSLPVPADILK 420
QY 421 LSLKRRSPROQFLGMRGSLTLIOLVOLFNDLSLKNDLGVIYLLIGDNNAKKVEEV 480
DB 421 LSLKRRSPROQFLGMRGSLTLIOLVOLFNDLSLKNDLGVIYLLIGDNNAKKVEEV 480
QY 481 LSVTPNDGPAKVHYGFIILKAQNKIAESIPYLKEGIESDPTGDDGRFYFHLGDMQVRGN 540
DB 481 LSVTPNDGPAKVHYGFIILKAQNKIAESIPYLKEGIESDPTGDDGRFYFHLGDMQVRGN 540
QY 541 KEAYKMYELGHRKGHPASVWQSLYNVGLKAPWMTKEGTGYTELVSLENNKILRDE 600
DB 541 KEAYKMYELGHRKGHPASVWQSLYNVGLKAPWMTKEGTGYTELVSLENNKILRDE 600
QY 601 GLAVVDKAKGLFPLPEDEMLREKGDWSOFTLMOQGRNENACKGAPKCTTLLEKPEPTGC 660
DB 601 GLAVVDKAKGLFPLPEDEMLREKGDWSOFTLMOQGRNENACKGAPKCTTLLEKPEPTGC 660
QY 661 RRGQIKYSIMPGTHVPHPTGPTNCLRMHLGLVLPKGGCKIRCANETRTWEKGVLIFD 720
DB 661 RRGQIKYSIMPGTHVPHPTGPTNCLRMHLGLVLPKGGCKIRCANETRTWEKGVLIFD 720
QY 721 DSFEHEWQDASSFLIFIVVWHPBELTPQQRSLPAI 758
DB 721 DSFEHEWQDASSFLIFIVVWHPBELTPQQRSLPAI 758

```

RESULT 2
 ABG72365
 ID ABG72365 standard; protein; 758 AA.
 XX
 XX ABG72365;
 AC
 XX
 DT 06-FEB-2003 (first entry)
 XX
 DE Human aspartyl (asparaginy1) beta-hydroxylase, HAAH.
 XX
 KW Human; enzyme; aspartyl (asparaginy1) beta-hydroxylase; HAAH; cytostatic;

KW immunostimulant; antibody; neoplasm; tumour; FB50; 86A; 5C7; 19B;
 KW brain tumour; glioma; glioblastoma; astrocytoma; haemangioma;
 KW pancreatic carcinoma; colon cancer; breast cancer; pancreatic cancer;
 KW liver cancer; cancer of the bile ducts; primary malignant CNS neoplasm;
 KW metastatic CNS neoplasm.
 XX
 XX Homo sapiens.
 OS
 XX
 Key
 FT Binding-site
 FT /label= "Antibody binding site"
 FT /note= "this region is specifically claimed in claim 36"
 FT 573..579
 FT /label= "Antibody binding site"
 FT /note= "this region is specifically claimed in claim 37"
 FT 600..700
 FT /label= "Catalytic domain 1"
 FT /note= "this region is specifically claimed in claim 7"
 FT 613..620
 FT /label= "Antibody binding site"
 FT /note= "this region is specifically claimed in claim 38"
 FT 650..700
 FT /label= "Catalytic domain 2"
 FT /note= "this region is specifically claimed in claim 24"
 FT 660..758
 FT /note= "alpha ketoglutarate binding site. This region is
 FT specifically claimed in claim 32"
 FT 679..697
 FT /note= "Epidermal growth factor-like domain; This region
 FT is specifically claimed in claim 33"
 XX
 XX US2002110559-A1.
 XX
 XX 15-AUG-2002.
 XX
 XX 17-MAY-2001; 2001US-00859604.
 XX
 XX 08-NOV-1999; 99US-00436184.
 XX
 XX (WAND/) WANDS J R.
 XX (DWON/) DE LA MONTE S M.
 XX (DEUT/) DEUTCH A H.
 XX (GHAN/) GHANBARI H A.
 PI Wands JR, De La Monte SM, Deutch AH, Ghanbari HA;
 XX
 XX WP1; 2003-066676/06.
 DR N-PSDB; ABS57868.
 XX
 PT Diagnosing neoplasm and inhibiting tumor growth, by contacting tissue of
 PT mammal with detectably-labeled antibody which binds to human aspartyl
 PT (asparaginy1) beta-hydroxylase.
 PT
 XX
 XX Claim 14; Page 3; 34pp; English.
 PS
 XX
 CC The invention relates to diagnosing a neoplasm and inhibiting tumour
 CC growth in a mammal, using an antibody that binds to human aspartyl
 CC (asparaginy1) beta-hydroxylase (HAAH). Diagnosing a neoplasm comprises
 CC contacting a tissue with a detectably-labelled antibody where an increase
 CC in level of antibody binding at tissue site compared to the level of
 CC binding to normal non-neoplastic tissue indicates the presence of a
 CC neoplasm at the tissue site. Inhibiting tumour growth in mammal involves
 CC administering the antibody conjugated to a cytotoxic agent to a mammal.
 CC Also included are a method of conferring an immune response to a mammal
 CC cell in a mammal, by administering the antibody, a method of inducing an
 CC HAAH-specific immune response in a mammal, by administering to the mammal
 CC an HAAH polypeptide (or a polynucleotide composition encoding the
 CC polypeptide, or its degenerate variant), a fragment of HAAH comprising an
 CC extracellular domain and lacking a cytoplasmic domain of HAAH, an
 CC antibody or its fragment that binds to HAAH (where the antibody is FB50,
 CC 86A, 5C7 or 19B), a hybridoma cell line chosen from hybridoma FB50,
 CC HA86A, HA15CA and HA219B, and a fragment of HAAH which lacks enzymatic
 CC activity or alpha-ketoglutarate binding domain and epidermal growth

CC factor (EGF)-like domain. The methods are useful for diagnosing neoplasm
 CC in a mammal, inhibiting tumour growth in a mammal, conferring an immune
 CC response to a brain tumour cell (e.g. glioma, glioblastoma, astrocytoma
 CC or haemangioma) in a mammal, for conferring immune response to a
 CC pancreatic carcinoma cell and for inducing a HAAH-specific immune
 CC response in a mammal. The method is useful for diagnosing malignant
 CC neoplasms derived from endodermal tissue, e.g. colon cancer, breast
 CC cancer, pancreatic cancer, liver cancer and cancer of the bile ducts. The
 CC method is also useful for diagnosing neoplasms of central nervous system
 CC (CNS) e.g. primary malignant CNS neoplasms of both neuronal and glial
 CC cell origin and metastatic CNS neoplasms, and for diagnosing brain
 CC tumours e.g. glioma, glioblastoma, astrocytoma or haemangioma. The
 CC present sequence represents human HAAH

XX Sequence 758 AA;

Query Match 100.0%; Score 4022; DB 6; Length 758;

Best Local Similarity 100.0%; Pred. No. 8.1e-111;

Matches 758; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAQRKAKSSGSSSSSSGSGSTAGSSSGARRETHGHKGRKGLSGTSFTFMV 60
 DB 1 MAQRKAKSSGSSSSGSGSTAGSSSGARRETHGHKGRKGLSGTSFTFMV 60
 QY 61 IALLGWTSAVAVVFPDLDVDEEVLGKLGIDADGDGFVDVDAKVLGLKERSTSEBAPV 120
 DB 61 IALLGWTSAVAVVFPDLDVDEEVLGKLGIDADGDGFVDVDAKVLGLKERSTSEBAPV 120
 QY 121 PEEAPEPTPEEQPVYAEFQNIIEDEAKEOIQSLHEHVAEHVEGSDLOQEDGPTPEPQ 180
 DB 121 PEEAPEPTPEEQPVYAEFQNIIEDEAKEOIQSLHEHVAEHVEGSDLOQEDGPTPEPQ 180
 QY 181 QEDDEFTMATDVDRFETLEPEVSHETESHVVEETVSODCNQDMEMMSSEONPDSSSE 240
 DB 181 QEDDEFTMATDVDRFETLEPEVSHETESHVVEETVSODCNQDMEMMSSEONPDSSSE 240
 QY 241 PVEDEELHHDPTDVTQVVEEOAVPELENEGEIETEVAPPEDNVEESQVYVEVSI 300
 DB 241 PVEDEELHHDPTDVTQVVEEOAVPELENEGEIETEVAPPEDNVEESQVYVEVSI 300
 QY 301 PVEEEOEVEPETNRTKTDEPEOKAVKVKKKPKLNEFDKTIKELDAEKLRRKGLIEEA 360
 DB 301 PVEEEOEVEPETNRTKTDEPEOKAVKVKKKPKLNEFDKTIKELDAEKLRRKGLIEEA 360
 QY 361 VNAFKELVRYKYPQSPARAYGKAQCEDDLAEKRSNEVLRGAIETTYQVVASLPVPADLLK 420
 DB 361 VNAFKELVRYKYPQSPARAYGKAQCEDDLAEKRSNEVLRGAIETTYQVVASLPVPADLLK 420
 QY 421 LSLKRSDROQFLGHWGSLTLQRLVQLFPNDTSLKNDLGVGVLIGDNDNAKRYEEV 480
 DB 421 LSLKRSDROQFLGHWGSLTLQRLVQLFPNDTSLKNDLGVGVLIGDNDNAKRYEEV 480
 QY 481 LSVTPNDGFAKVHYGFLKAQNKLAESIPLYLKEGIESGDEGTDGFFYFHLGAMORVAN 540
 DB 481 LSVTPNDGFAKVHYGFLKAQNKLAESIPLYLKEGIESGDEGTDGFFYFHLGAMORVAN 540
 QY 541 KEAYKVELGHKRGHPASVWORSILYNNGLKAQPMWPKETGYTELKYSLERWKLIRDE 600
 DB 541 KEAYKVELGHKRGHPASVWORSILYNNGLKAQPMWPKETGYTELKYSLERWKLIRDE 600
 QY 601 GLAWMDAKGLFLPEDENLRKGGWQFTLMQOGRRENKCKGAPKCTTLEKPEPTTG 660
 DB 601 GLAWMDAKGLFLPEDENLRKGGWQFTLMQOGRRENKCKGAPKCTTLEKPEPTTG 660
 QY 661 RRGQIKYSIMHPTGPTNCRMLGLVTPREGCKIRCANEETRWEEGKVLIFD 720
 DB 661 RRGQIKYSIMHPTGPTNCRMLGLVTPREGCKIRCANEETRWEEGKVLIFD 720
 QY 721 DSFEHEWQDASSFRLLFIYDVVHPELTPOORSLPAI 758
 DB 721 DSFEHEWQDASSFRLLFIYDVVHPELTPOORSLPAI 758

RESULT 3

ID ADA00639 standard; protein; 758 AA.

AC ADA00639;

DT 06-NOV-2003 (first entry)

DE Human aspartyl (asparaginyl) beta-hydroxylase (HAAH).

KW Tumour growth inhibition; human aspartyl (asparaginyl) beta-hydroxylase;

KW HAAH hydroxylation; NOTCH polypeptide;

KW epidermal growth factor-like repeat; EGF-like repeat; tumour cell;

KW malignant neoplasm; colon cancer; breast cancer; pancreatic cancer;

KW liver cancer; cancer of the bile duct; cancer the central nervous system;

KW CNS; cytosolic; enzyme; human.

OS Homo sapiens.

PN US2003031670-A1.

PD 13-FEB-2003.

PF 08-NOV-1999; 99US-00436184.

PR 08-NOV-1999; 99US-00436184.

PA (WAND/) WANDS J R.

PA (DMON/) DE LA MONTE S M.

PA (INCE/) INCE N.

PA (CARL/) CARLSON R I.

PT Wands JR, De La Monte SM, Ince N, Carlson RI;

DR WPI; 2003-605701/57.

DR N-PSDB; ADA00640.

PT Inhibiting tumor growth or killing tumor cells (e.g. cancer of the colon,

PT breast, pancreatic, liver or the central nervous system), by

PT administering an inhibitor of the human aspartyl (asparaginyl) beta-

PT hydroxylase.

PS Disclosure; Page 2; 30pp; English.

CC The present invention relates to a method for inhibiting tumour growth in
 CC a mammal. The method comprises administering to the mammal a compound,
 CC which inhibits the expression or enzymatic activity of a human aspartyl
 CC (asparaginyl) beta-hydroxylase (HAAH). The compound may inhibit HAAH
 CC hydroxylation of a NOTCH polypeptide. In particular, the compound may
 CC inhibit hydroxylation of an epidermal growth factor (EGF)-like repeat
 CC sequence in a NOTCH polypeptide. The methods are useful for inhibiting
 CC tumour growth or killing tumour cells, or for diagnosing or
 CC prognosticating a malignant neoplasm. In particular, the tumour or
 CC neoplasm is colon cancer, breast cancer, pancreatic cancer, liver cancer,
 CC cancer of the bile ducts, or cancer or tumour of the central nervous
 CC system (CNS). The present sequence represents HAAH.

XX Sequence 758 AA;

Query Match 100.0%; Score 4022; DB 6; Length 758;

Best Local Similarity 100.0%; Pred. No. 8.1e-111;

Matches 758; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAQRKAKSSGSSSSGSGSTAGSSSGARRETHGHKGRKGLSGTSFTFMV 60
 DB 1 MAQRKAKSSGSSSSGSGSTAGSSSGARRETHGHKGRKGLSGTSFTFMV 60
 QY 61 IALLGWTSAVAVVFPDLDVDEEVLGKLGIDADGDGFVDVDAKVLGLKERSTSEBAPV 120
 DB 61 IALLGWTSAVAVVFPDLDVDEEVLGKLGIDADGDGFVDVDAKVLGLKERSTSEBAPV 120
 QY 121 PEEAPEPTPEEQPVYAEFQNIIEDEAKEOIQSLHEHVAEHVEGSDLOQEDGPTPEPQ 180
 DB 121 PEEAPEPTPEEQPVYAEFQNIIEDEAKEOIQSLHEHVAEHVEGSDLOQEDGPTPEPQ 180


```

QY 179 PQQEDDEFMAATDVDDREFETLEPEVSHHETSHYHVEETVSQDCNQDMENMSEQENDS 238
DB 121 PQQEDDEFMAATDVDDREFETLEPEVSHHETSHYHVEETVSQDCNQDMENMSEQENDS 180
QY 239 SEPVEDERLHHDDVDTVYQVVEQAVYEPLENIGIETITETVAPPEPNVPEDSQVYVEV 298
DB 181 SEPVEDERLHHDDVDTVYQVVEQAVYEPLENIGIETITETVAPPEPNVPEDSQVYVEV 240
QY 299 SIFPVEEQOEVPPEPT 313
DB 241 SIFPVEEQOEVPPEPT 255

RESULT 7
ABU69516
ID ABU69516 standard; protein; 255 AA.
XX
AC ABU69516;
XX
DT 05-JUN-2003 (first entry)
XX
DE Human lung cancer-associated protein L979P.
XX
KM Human; lung cancer; lung tumour; cytostatic; vaccine; T cell expansion;
CD4; CD8.
XX
OS Homo sapiens.
XX
PN US2002197669-A1.
XX
PD 26-DEC-2002.
XX
PF 03-MAY-2001; 2001US-00849626.
XX
PR 13-DEC-2000; 2000US-00736457.
XX
PA (BANG/) BANGUR C S.
PA (FANG/) FANGER G R.
PA (WANG/) WANG A.
PA (WANG/) WANG T.
PA (SWIT/) SWITZER A P.
PA (MCNE/) MCNEILL P D.
PA (CLAP/) CLAPPER J D.
XX
PI Banour CS, Fanger GR, Wang A, Wang T, Switzer AP, McNeill PD;
PI Clapper JD;
XX
DR WPI; 2003-352750/33.
DR N-PSDB; ACA12072.
XX
PT Novel lung cancer polynucleotide encoding lung cancer protein, useful for
PT detecting the presence of lung cancer in a patient, and in pharmaceutical
PT compositions, e.g. vaccines, for treating lung cancer.
XX
PS Example 5; Page; 72pp; English.
XX
XX The invention relates to a polynucleotide encoding a lung tumour protein,
XX comprising a sequence selected from any of the 14 sequences mentioned in
XX the specification, or a sequence (S2) mentioned in specification,
XX complement of S1, sequences consisting of at least 20 contiguous residues
XX of S1, sequences that hybridise to S1, sequences having 75%, preferably
XX 90%, identity to S1, or degenerate variants of S1. Also included are an
XX isolated polypeptide (comprising a sequence (S3) selected from any one of
XX the 4 amino acid sequences mentioned in the specification, a sequence
XX encoded by the polynucleotide, or sequences having at least 70%,
XX preferably 90%, identity to a sequence encoded by the polynucleotide), an
XX expression vector comprising the polynucleotide operably linked to an
XX expression control sequence, a host cell transformed or transfected with
XX the vector, an isolated antibody (or its antigen-binding fragment) that
XX specifically binds to the polypeptide, detecting the presence of a cancer
XX in a patient, a fusion protein comprising the polypeptide, an
XX oligonucleotide that hybridises to S1 under moderately stringent
XX conditions, stimulating and/or expanding T cells specific for a tumour

```

```

CC protein (comprising contacting T cells with the polynucleotide, protein
CC or antigen-presenting cells, under conditions and for a time sufficient
CC to permit the stimulation and/or expansion of T cells) and inhibiting the
CC development of a cancer in a patient (by incubating CD4+ and/or CD8+ T
CC cells isolated from a patient with the polynucleotide, protein or antigen
CC presenting cells that express the polynucleotide, such that T cells
CC proliferate, administering to the patient an effective amount of the
CC proliferated T cells, and thus inhibiting the development of a cancer in
CC the patient. The polynucleotide, protein and cells are useful in a
CC composition for stimulating an immune response in a patient, and for
CC treating a cancer in a patient (particularly lung cancer). The
CC oligonucleotide is useful for determining the presence of a cancer in a
CC patient. The protein and oligonucleotides are useful in pharmaceutical
CC compositions, e.g. vaccines. The polynucleotide is also useful as a probe
CC or primer for nucleic acid hybridisation, and in the design and
CC preparation of ribozyme molecules for inhibiting expression of tumour
CC polypeptides and proteins in tumour cells. An amplified portion of the
CC polynucleotide is useful for isolating a full-length gene from a suitable
CC library. The present sequence is a protein encoded by a cDNA (full
CC length, extended or partial) isolated from a library derived from lung
CC tumour/ cancer cells. Note: The sequence data for this patent did not
CC form part of the printed specification, but was obtained in electronic
CC format directly from the USPTO at
CC seqdata.uspto.gov/sequence.html?docid=20020197669
XX
SQ Sequence 255 AA:
XX
Query Match 32.8%; Score 1320; DB 6; Length 255;
Best Local Similarity 98.8%; Pred. No. 1.9e-96;
Matches 252; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 59 MVIALLGWTSVAWVDFDLVYEEVLGKLGTYDADGDDPVDPAKVLGKERSTSPA 118
DB 1 MVIALLGWTSVAWVDFDLVYEEVLGKLGTYDADGDDPVDPAKVLGKERSTSPA 60
QY 119 VPPEABPHTEPEEQVPEABPQNIIEBAKQIQLHEMVAHVEGEDLQQEDGPTGE 178
DB 61 VPPEABPHTEPEEQVPEABPQNIIEBAKQIQLHEMVAHVEGEDLQQEDGPTGE 120
QY 179 PQQEDDEFMAATDVDDREFETLEPEVSHHETSHYHVEETVSQDCNQDMENMSEQENDS 238
DB 121 PQQEDDEFMAATDVDDREFETLEPEVSHHETSHYHVEETVSQDCNQDMENMSEQENDS 180
QY 239 SEPVEDERLHHDDVDTVYQVVEQAVYEPLENIGIETITETVAPPEPNVPEDSQVYVEV 298
DB 181 SEPVEDERLHHDDVDTVYQVVEQAVYEPLENIGIETITETVAPPEPNVPEDSQVYVEV 240
QY 299 SIFPVEEQOEVPPEPT 313
DB 241 SIFPVEEQOEVPPEPT 255

RESULT 8
ABU66419
ID ABU66419 standard; protein; 255 AA.
XX
AC ABU66419;
XX
DT 22-MAY-2003 (first entry)
XX
DE Lung cancer therapy and diagnosis associated protein #43.
XX
DE Lung cancer; cytostatic; vaccine; gene therapy; cancer.
XX
KM Lung cancer; cytostatic; vaccine; gene therapy; cancer.
XX
OS Homo sapiens.
XX
PN US2002172952-A1.
XX
PD 21-NOV-2002.
XX
PF 10-JUL-2001; 2001US-00902941.
XX
PR 30-JUN-1999; 99US-00346492.

```

PR 15-OCT-1999; 99US-00419356.
PR 17-DEC-1999; 99US-00468687.
PR 30-DEC-1999; 99US-00476300.
PR 06-MAR-2000; 2000US-00519642.
PR 22-MAR-2000; 2000US-00533077.
PR 10-APR-2000; 2000US-00546259.
PR 27-APR-2000; 2000US-00560406.
PR 05-JUN-2000; 2000US-00589184.
PR 11-JUL-2000; 2000US-00614124.
PR 29-AUG-2000; 2000US-00651563.
PR 08-SEP-2000; 2000US-00658824.
PR 26-SEP-2000; 2000US-00671325.
PR 06-OCT-2000; 2000US-00677419.
PR 30-OCT-2000; 2000US-00702705.
PR 13-DEC-2000; 2000US-00736457.
PR 03-MAY-2001; 2001US-00849626.

XX (CORI-) CORIXA CORP.

PI Henderson RA, Wang T, Matanabe Y, Johnson JC, Retter MW;
PI Durham M, Carter D, Fanger GR, Vedvick TS, Bangur CS, McNabb A;

DR WPI; 2003-328427/31.

PT New polynucleotide, useful for preparing a composition for treating or
PT inhibiting development of cancer, e.g. lung cancer.

PS Example 5; SEQID NO 1807; 82pp; English.

XX The invention describes an isolated polynucleotide comprising one of 32
CC sequences, complement or degenerate variants of them. The polynucleotide
CC is useful for preparing a composition e.g. a vaccine or for gene therapy.
CC for treating or inhibiting development of cancer, e.g. lung cancer. This
CC sequence represents a polypeptide associated with the compositions and
CC methods for the therapy and diagnosis of lung cancer
XX

SQ Sequence 255 AA;

Query Match 33.8%; Score 1320; DB 6; Length 255;

Best Local Similarity 96.8%; Pred. No. 1.9e-96;
Matches 252; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 59 MVALLGVTSTVAWVDFLDVYBEVLGKLGIDADGDPDPAKYLGLKERSTSEPA 118
DB 1 MVALLGVTSTVAWVDFLDVYBEVLGKLGIDADGDPDPAKYLGLKERSTSEPA 60
QY 119 VPPEASPTPEPEQVVEAEPQNIIEDEAKQIQSLHEMVHAHVEGEDLQGEPTGE 178
DB 61 VPPEASPTPEPEQVVEAEPQNIIEDEAKQIQSLHEMVHAHVEGEDLQGEPTGE 120
QY 179 PQQDDDFLMAVDVDFEFTLEPSVSHETESHVVEETVSQONOMEEMMSQENPDS 238
DB 121 PQQDDDFLMAVDVDFEFTLEPSVSHETESHVVEETVSQONOMEEMMSQENPDS 180
QY 239 SEPVEBERLHNDTDVLYQVVEEQAVYEPLENIGIEITETVAPPEPNPYEDSQVIEEV 298
DB 181 SEPVEBERLHNDTDVLYQVVEEQAVYEPLENIGIEITETVAPPEPNPYEDSQVIEEV 240
QY 299 SIFPVEEQVEVPPT 313
DB 241 SIFPVEEQVEVPPT 255

RESULT 9

ABB61986
ID ABB61986 standard; protein; 422 AA.

AC ABB61986;

DT 26-MAR-2002 (first entry)

XX Drosophila melanogaster polypeptide SEQ ID NO 12750.
XX

KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.

OS Drosophila melanogaster.

PN WO200171042-A2.

XX 27-SEP-2001.

PF 23-MAR-2001; 2001WO-US009231.

PR 23-MAR-2000; 2000US-0191637P.

PR 11-JUL-2000; 2000US-00614150.

XX (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EM;

DR WPI; 2001-656860/75.

DR N-PADB; ABL06089.

PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signaling and cell-cell
PT interactions.

PS Disclosure; SEQ ID NO 12750; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signaling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA
CC sequences (AB101840-AB16175) and the encoded proteins (AB57737-
CC AB572072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WPI at ftp.wpi.int/pub/published_pct_sequences
XX

SQ Sequence 422 AA;

Query Match 23.2%; Score 933; DB 4; Length 422;

Best Local Similarity 45.0%; Pred. No. 2.5e-65;
Matches 184; Conservative 75; Mismatches 144; Indels 6; Gaps 5;

QY 355 GKLEEVNAKKELVKRYKPSPRARYGKQCEDDLAEKRSNEVLRGALIEYQVASLPDV 414
DB 15 GNYAQLRSFNTLTTFNFAHPSAHLGPARLELJLAKKERENQRLWEALDAVKRYLAAGEL 74
QY 415 PADLLKLSLRRS--DROCFIAGMRGSLTLQRLVQLFPNDTSIKNDLGVGYLLIGDND 472
DB 75 VASNOEFGTAGESCIENLRFLGHHRQATTHELINRLRPDRRLRNQSLTYLVMNNLQQ 134
QY 473 AKRYVEVLSTVNDGPAKVHYGFILKA-ONKLAESIPLYKGEIGSDPGTDDGRFTFHL 531
DB 135 VERVAAVETLLVMNNNAVAQIHYGIALRQFADYAKALPYLKYVVESEEGTQEAFFYLSL 194
QY 532 GDMQOVGNK-EAYKMYEIGHKRGHPASVQBSLYVNVGKAPQWTPKCTGYTELKSL 590
DB 195 GETWQRLSKSEALVYGVKAVAGFPASLTQRSINIEPRIRAPQFPQPKETGYRQLEKL 254
QY 591 ERNWKLRIDEGLVAMDKAKGLFLPEDENLREKQDWSQFTLMOQGRNBNACKAPKCTTL 650
DB 255 TLNMRALRDRGLALNLSG-SGFDEAELDLKDGWQGYELYYAQGRVYKDCNRAPITCSL 313
QY 651 LEKFPETTCGRGQIKSYNHPGTHVPHPTNCRIRMLGLVIER-EGCKTRCANETR 709
DB 314 LEEFPESAGCRQKQVSVQAKTHVPHGPTNCRIRALHTLAAPPERASLVAEOER 373
QY 710 TWEEGKVLFPDPSPEHEWQDASFRLLFIVDVVWHPFLTPQGRSLPAI 758
DB 374 TWEEGELFIPDPSPEHEWQDASFRLLFIVDVVWHPFLTPQGRSLPAI 422

RESULT 10
 AAB73682
 ID AAB73682 standard; protein; 369 AA.
 XX
 AC AAB73682;
 XX
 DT 11-SEP-2001 (first entry)
 XX
 DE Human oxidoreductase protein ORP-15.
 XX
 KW Human oxidoreductase protein; ORP; cell proliferative disorder;
 KW arteriosclerosis; cirrhosis; psoriasis; cancer; endocrine disorder;
 KW diabetes mellitus; diabetes insipidus; dwarfism; hirsutism; amenorrhoea;
 KW osteoporosis; metabolic disorder; obesity; phenylketonuria;
 KW hypercholesterolaemia; reproductive disorder; infertility;
 KW ovulatory defect; menstrual cycle defect; endometriosis; chromosome 22;
 KW polycystic ovary disease; spermatogenesis disruption; impotence;
 KW neurological disorder; epilepsy; stroke; Alzheimer's disease;
 KW Huntington's disease; Parkinson's disease; Creutzfeldt-Jakob disease;
 KW meningitis; cerebral palsy; muscular dystrophy; mood disorder; anxiety;
 KW schizophrenia disorder; infection; autoimmune disorder;
 KW inflammatory disorder; acquired immunodeficiency syndrome; AIDS; asthma;
 KW allergy; Crohn's disease; atopic dermatitis; gout; multiple sclerosis;
 KW rheumatoid arthritis; ulcerative colitis; drug screening;
 KW toxicity screening; transgenic animal; SNP detection; gene therapy; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200144448-A2.
 XX
 PD 21-JUN-2001.
 XX
 PF 07-DEC-2000; 2000WO-US033158.
 XX
 PR 16-DEC-1999; 99US-0172367P.
 XX
 PA (INCY-) INCYTE GENOMICS INC.
 XX
 PI Yue H, Lal P, Tang YT, Hillman JL, Baughn MR, Azimzai Y, Lu DM;
 XX
 DR MPI; 2001-390245/41.
 XX
 DR N-PSDB; AAB72437.
 XX
 PT Novel human oxidoreductase protein (ORP) useful for diagnosing, treating
 PT and preventing cell proliferative, neurological, viral, reproductive and
 PT autoimmune/inflammatory disorders associated with abnormal expression of
 PT ORP.
 XX
 PS Claim 1; Page 110-111; 136pp; English.
 XX
 CC Sequences AAB73668-AAB73694 represent 27 novel human oxidoreductase
 CC proteins, designated ORP-1 to ORP-27 respectively, and sequences AAB74223
 CC -AAB74249 represent cDNAs encoding ORP-1 to ORP-27. Human ORP proteins
 CC and nucleic acids are useful for diagnosing, treating or preventing cell
 CC proliferative disorders (e.g. arteriosclerosis, cirrhosis, psoriasis,
 CC cancers); endocrine disorders (e.g., Type I or II diabetes mellitus,
 CC diabetes insipidus, dwarfism, hirsutism, amenorrhoea, osteoporosis);
 CC metabolic disorders (e.g., obesity, phenylketonuria,
 CC hypercholesterolaemia); reproductive disorders (e.g., infertility,
 CC ovulatory and menstrual cycle defects, endometriosis, polycystic ovary
 CC disease, disruption of spermatogenesis, impotence); neurological
 CC disorders (e.g., epilepsy, stroke, Alzheimer's disease, Huntington's
 CC disease, Parkinson's disease, meningitis, Creutzfeldt-Jakob disease,
 CC cerebral palsy, muscular dystrophy, mood, anxiety and schizophrenic
 CC disorders); viral, bacterial, fungal and parasitic infections; and
 CC autoimmune/inflammatory disorders such as acquired immunodeficiency
 CC syndrome (AIDS), allergies, asthma, Crohn's disease, atopic dermatitis,
 CC gout, multiple sclerosis, rheumatoid arthritis or ulcerative colitis.
 CC Human ORP proteins and nucleic acids can be used to identify compounds
 CC which modulate their activity or expression. ORP nucleic acid sequences
 CC may also be used for assessing the toxicity of a test compound, to detect
 CC upstream sequences such as promoters and regulatory elements, and to
 CC create knock out or knock in animals or transgenic animals to model human

CC disease. Oligonucleotide primers derived from ORP gene sequences may be
 CC used to detect single nucleotide polymorphisms (SNPs) and for mapping the
 CC naturally occurring genomic sequences. Antibodies specific for ORP
 CC proteins may be used in the diagnosis of disorders associated with
 CC aberrant ORP expression, in assays to monitor patients being treated with
 CC ORP or modulators thereof, and for assessing toxicity of potential drugs
 CC
 XX
 SQ Sequence 369 AA;
 XX
 Query Match 6.8%; Score 274.5; DB 4; Length 369;
 Best Local Similarity 30.1%; Pred. No. 5.4e-13;
 Matches 75; Conservative 44; Mismatches 105; Indels 25; Gaps 9;
 QY 526 RFVHLDDAMQVRNKNKAYKMYELG--HK-----RGHF-----ASYWQSLVNVNGAKQAP 575
 DB 125 KLYNNLDGVYAR-----YWSGGRJHKRGREGRYNRSPTQKVEVFLPDLPTTPY 178
 QY 576 WTPKETYTELIVKSLERNMKLIRDEGLAVMDKAKGLFPE--DENTLEKQPMQFTLMQO 633
 DB 179 FSRDAQHND--VEVLERNFQTLICEFTTIVKAFSNGSLPQGMKNNSPTSGEMFTFVLVNO 236
 QY 634 GRRENACKAPKCTTLEKPPETTCGR-RQIKYSLMHPEHTHWPTGPTNCLKRLNLG 692
 DB 237 GVCVPRNCRKCPRTYRLGLSLRTICGNVPGNACISVLSPTVTEHYGPTNIRICHLG 296
 QY 693 LVIPKSGCKIRCANETRTWEGKVLIPDPSFEHFWODASGF--RLIFIVDVWHELTLP 749
 DB 297 LKTP-NGCELVYGGEPQCMABRGRLPLDDSFHLAAHFEGSAEDQPRVYFWMDLWHPNVAA 355
 QY 750 QQRSLPAI 758
 DB 356 AERQALDPI 364
 XX
 RESULT 11
 AAB43327
 ID AAB43327 standard; protein; 236 AA.
 XX
 AC AAB43327;
 XX
 DT 08-FEB-2001 (first entry)
 XX
 DE Human ORFX ORF3091 polypeptide sequence SEQ ID NO:6182.
 XX
 KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
 KW vulnary; antiproliferative; antiparkinsonian; neurotropic; neuroprotective;
 KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiac;
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
 KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
 KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
 KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
 KW cholesterol ester storage; systemic lupus erythematosus; infection;
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
 KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
 KW thrombosis; contraceptive.
 XX
 OS Homo sapiens.
 XX
 PN WO200058473-A2.
 XX
 PD 05-OCT-2000.
 XX
 PF 31-MAR-2000; 2000WO-US008621.
 XX
 PR 31-MAR-1999; 99US-0127607P.
 XX
 PR 02-APR-1999; 99US-0127636P.
 XX
 PR 05-APR-1999; 99US-0127782P.
 XX
 PR 30-MAR-2000; 2000US-00540763.
 XX
 PA (CURA-) CURAGEN CORP.

XX Shinkets RA, Leach M;
XX WPI; 2000-602362/57.
DR N-PESDB; AAC77536.
XX
XX Novel nucleic acids and peptides derived from open reading frame X,
PT useful for treating e.g. cancers, proliferative disorder,
PT neurodegenerative disorders and cardiovascular disease.
XX
PS Claim 11; Page 5365; 5507pp; English.
XX
CC AAC74446 to AAC7606 encode the proteins given in AAB40237 to AAB43397,
CC which represent the human ORF open reading frames 1 to 3161. The ORF
CC sequences have activities such as: cytostatic; hepatotropic; vulnery;
CC antiproliferative; antiparkinsonian; nocotropic; neuroprotective; osteopathic;
CC anticonvulsant; antiarthritic; immunosuppressant; immunostimulant;
CC cardiatic; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive;
CC dermatological; immunosuppressive; antineoplastic; antibacterial;
CC antiviral; antifungal; antirheumatic; antihypertensive; and antianemic. The
CC sequences can be used for determining the presence of or predisposition
CC to, or preventing or treating pathological conditions associated with an
CC ORF-associated disorder. The nucleic acids can be used to express ORF
CC proteins in gene therapy vectors. The proteins and nucleic acids may be
CC used to treat cancers, proliferative disorders, neurodegenerative
CC disorders, osteoarthritis, graft vs host disease, cardiovascular disease,
CC diabetes mellitus, hypertension, hypothyroidism, cholesterol ester
CC storage, systemic lupus erythematosus, severe combined immunodeficiency
CC (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune
CC disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and
CC cartilage damage, nocturnal haemoglobinuria, antineoplastic disease, to
CC enhance coagulation, to inhibit thrombosis, and as a contraceptive
XX
SQ Sequence 236 AA;
Query Match 6.8%; Score 271.5; DB 3; Length 236;
Best Local Similarity 31.2%; Pred. No. 4.7e-13;
Matches 72; Conservative 40; Mismatches 100; Indels 19; Gaps 8;
QY 544 YKMYELG--HK-----RGHF-----ASVMORSLYNNGKAQPMWPKETGYTELKYSERN 593
DB 4 YSNMGGRHNGHIGTEQGRYLNRSRISQKPEYFPLDLPTTYFSRDAQKH--VEVLERN 61
QY 594 WKLRDEGLAVMDKAKGLFPE--DENLREKDWSOFTLMOQGRNENACKGAPKCTTL 651
DB 62 FQTLICEFETLYKAFSNCSLPGQWKNSTPGEWTFEVLVQGVCPNRCCKPRTYRL 121
QY 652 EKFEETTCGR-RGQIKYSIMHGRTHVWHTGTPTNCRMLMHLGLVTPKGGCKIRCANETRT 710
DB 122 GSLRTCIANNVFNACISVLSPGVITTEHYGPTNIRIRCHGLKTP--NGCELVVGGEPOC 180
QY 711 WEEGKVLIFDDSFHEHVMQDASSF---RLIFIVDMHPELTPOQRSLPAI 758
DB 181 WABRCILFDYSYQHAEVYNSDKRIRICLIMDIWHPNLSFAERAEALVLI 251
RESULT 12
AAB68322
ID AAB68322 standard; protein, 264 AA.
XX
AC AAB68322;
XX
XX 20-NOV-2003 (first entry)
DE Photorhabdus luminescens protein sequence #1419.
XX
XX Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;
KM detection; food; gene expression; plant; animal; microorganism; toxin;
KM antibiotic; bioplastic; virulence factor; disease model; plague;
XX whooping cough.
XX OS Photorhabdus luminescens.
XX

PN WO200294867-A2.
XX
XX 28-NOV-2002.
PD
XX
XX 07-FEB-2002; 2002WO-IB003040.
PF
XX
XX 07-FEB-2001; 2001FR-00001659.
PR
XX
XX (INSP) INST PASTEUR.
PA (CNRS) CNRS CENT NAT RECH SCT.
XX
XX Duchaud E, Taourit S, Glaser P, Frangeul L, Kunst F, Danchin A;
PI Buchrieser C;
XX WPI; 2003-148459/14.
XX
XX Genomic sequence of Photorhabdus luminescens and encoded polypeptides,
PT useful e.g. as therapeutic antimicrobials and agricultural pesticides.
XX
PS Claim 2; SEQ ID NO 1419; 1205pp; French.
XX
CC The invention relates to the isolation of genes and their encoded
CC proteins from Photorhabdus luminescens. The isolated sequences are
CC sources of probes and primers for detecting the genome of P. luminescens
CC and related species; to study polymorphisms; for gene analysis and for
CC detection/amplification of the genes. Antibodies (Ab) raised against the
CC polypeptides encoded by the genes are used for detection/identification
CC of P. luminescens, e.g. in foods. The genes, proteins, Ab and cells that
CC carry a gene-containing vector are used to select compounds that
CC modulate, regulate, induce or inhibit expression of the genes in plants,
CC animals or microorganisms other than P. luminescens and are able to alter
CC response or sensitivity to toxins and antibiotics produced by P.
CC luminescens. Cells transformed to express the genes are useful for
CC recombinant production of the proteins, particularly toxins and
CC antibacterials useful as insecticides, bactericides and fungicides. The
CC genes, proteins, vectors containing the genes and Ab are also useful
CC therapeutically (to treat microbial infection by bacteria or fungi that
CC are sensitive to P. luminescens-encoded toxins or antibiotics) and as
CC biopesticides. Other uses of the genes and the proteins are as virulence
CC factors and for identifying targets of human diseases for which P.
CC luminescens is a model (particularly plague and whooping cough). This
CC sequence represents one of the isolated P. luminescens proteins
XX
SQ Sequence 264 AA;
Query Match 6.5%; Score 262; DB 6; Length 264;
Best Local Similarity 29.1%; Pred. No. 3.2e-12;
Matches 66; Conservative 39; Mismatches 114; Indels 8; Gaps 3;
QY 533 DAMQRYNKEAYKMYELGHRKGFASVWORSLYNNGKAQPMW-TPKETGYTELKYSLE 591
DB 35 DCLKRVED---MRYPLQQR---APQADAKYIMPLSTPMDITNSFPOLQPLVTSIM 87
QY 592 RNMKLTIDEGGLAVMDKAKGLFPEDENLRRKGMWSQFTLMOQGRNENACKGAPKCTTL 651
DB 88 NNADKIKQELNAAVYSGESQYITTYEHYLGQKQWKALYLFRNQPNNAAVANIILPATWHTF 147
QY 652 EKFEETTCGRGQIKYSIMHGRTHVWHTGTPTNCRMLMHLGLVTPKGGCKIRCANETRTW 711
DB 148 NNELRDMHCEFLLEHVEVLDPGTVIKHCHLMFTLMLHRAVDLPASHCELIIVNEARCW 207
QY 712 EEGKVLIFDDSFHEHVMQDASSFRLIFIVDMHPELTPOQRSLPAI 758
DB 208 KEGRCILFDYSYQHAEVYNSDKRIRICLIMDIWHPNLSFAERAEALVLI 254
RESULT 13
AAU29679
ID AAU29679 standard; protein, 104 AA.
XX
XX AAU29679;
XX
XX 18-DEC-2001 (first entry)
DT

```
XX  Novel human secreted protein #170.
DE
XX
XX  Human; vaccination; gene therapy; nutritional supplement;
KM  stem cell proliferation; haematopoiesis; nerve tissue regeneration;
KM  immune suppression; immune stimulation; anti-inflammatory; leukaemia.
XX
OS  Homo sapiens.
XX
XX  WO200179449-A2.
XX
XX  25-OCT-2001.
XX
XX  16-APR-2001; 2001WO-US0086656.
XX
XX  18-APR-2000; 2000US-00552929.
XX  26-JAN-2001; 2001US-00770160.
XX
XX  (HYSE-) HYSEQ INC.
XX
XX  Tang YT, Liu C, Drmanac RT;
PI  WPI; 2001-611725/70.
XX
XX  Nucleic acids encoding a range of human polypeptides, useful in genetic
PT  vaccination, testing and therapy.
XX
XX  Claim 20; Page 179; 765pp; English.
XX
XX  The invention relates to novel human secreted polypeptides. The
CC  polypeptides and antibodies to the polypeptides are useful for
CC  determining the presence of or predisposition to a disease associated
CC  with altered levels of polypeptide. The polypeptides are also useful for
CC  identifying agents (agonists and antagonists) that bind to them. Cells
CC  expressing the proteins are useful for identifying a therapeutic agent
CC  for use in treatment of a pathology related to aberrant expression or
CC  physiological interactions of the polypeptide. Vectors comprising the
CC  nucleic acids encoding the polypeptides and cells genetically engineered
CC  to express them are also useful for producing the proteins. The proteins
CC  are useful in genetic vaccination, testing and therapy, and can be used
CC  as nutritional supplements. They may be used to increase stem cell
CC  proliferation; to regulate haematopoiesis; and in bone, cartilage, tendon
CC  and/or nerve tissue growth or regeneration; immune suppression and/or
CC  stimulation; as anti-inflammatory agents; and in treatment of leukaemias.
CC  AAU9510-AAU3304 represent the amino acid sequences of novel human
CC  secreted proteins of the invention
XX
XX  Sequence 104 AA;
SQ
XX
XX  Query Match          6.0%; Score 241; DB 4; Length 104;
XX  Best Local Similarity 54.7%; Pred. No. 3.7e-11;
XX  Matches 47; Conservative 11; Mismatches 26; Indels 0; Gaps 0;
XX
XX  13 SSSSGSGSGTSGAGSSPGARRRKHGHNKRGKGLSTGTFWMTALIGWTSYAV 72
XX  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX  13 SFTGTGCSSTVRAAAPGTPQRPSPDAMTKRGKLSYFTSMWTALIGWTSYAV 72
XX
XX  73 VWFEDLVDFEVLGKIGIYDADGDP 98
XX  ||||| : : : : : : : : : : : : : : : : : : : : : :
XX  73 VWFEDLVDFEVLGKIGIYDADGDP 98
XX
XX  RESULT 14
XX  ABU21099
XX  ID ABU21099 standard; protein; 299 AA.
XX
XX  ABU21099;
XX
XX  19-JUN-2003 (first entry)
XX
XX  Protein encoded by Prokaryotic essential gene #6626.
XX
XX  Antisense; prokaryotic essential gene; cell proliferation; drug design.
```

```
XX  Burkholderia fungorum.
OS
XX
XX  WO200277183-A2.
XX
XX  03-OCT-2002.
XX
XX  21-MAR-2002; 2002WO-US009107.
XX
XX  21-MAR-2001; 2001US-00815242.
XX  06-SEP-2001; 2001US-00948993.
XX  25-OCT-2001; 2001US-0342923P.
XX  08-FEB-2002; 2002US-00072851.
XX  06-MAR-2002; 2002US-0362699P.
XX
XX  (ELIT-) ELITRA PHARM INC.
XX
XX  Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
PI  Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX  WPI; 2003-029926/02.
XX  N-PSDB; ACPA24969.
XX
XX  New antisense nucleic acids, useful for identifying proteins or screening
PT  for homologous nucleic acids required for cellular proliferation to
PT  isolate candidate molecules for rational drug discovery programs.
XX
XX  Claim 25; SEQ ID NO 49023; 1766pp; English.
XX
XX  The invention relates to an isolated nucleic acid comprising any one of
CC  the 6213 antisense sequences given in the specification where expression
CC  of the nucleic acid inhibits proliferation of a cell. Also included are:
CC  (1) a vector comprising a promoter operably linked to the nucleic acid
CC  encoding a polypeptide whose expression is inhibited by the antisense
CC  nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC  polypeptide or its fragment whose expression is inhibited by the
CC  antisense nucleic acid; (4) an antibody capable of specifically binding
CC  the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC  proliferation or the activity of a gene in an operon required for
CC  proliferation; (7) identifying a compound that influences the activity of
CC  the gene product or that has an activity against a biological pathway
CC  required for proliferation, or that inhibits cellular proliferation; (8)
CC  identifying a gene required for cellular proliferation or the biological
CC  pathway in which a proliferation-regulated gene or its gene product lies
CC  or a gene on which the test compound that inhibits proliferation of an
CC  organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC  compound's activity; (11) a culture comprising strains in which the gene
CC  product is overexpressed or underexpressed; (12) determining the extent
CC  to which each of the strains is present in a culture or collection of
CC  strains; or (13) identifying the target of a compound that inhibits the
CC  proliferation of an organism. The antisense nucleic acids are useful for
CC  identifying proteins or screening for homologous nucleic acids required
CC  for cellular proliferation to isolate candidate molecules for rational
CC  drug discovery programs, or for screening homologous nucleic acids
CC  required for proliferation in cells other than S. aureus, S. typhimurium,
CC  K. pneumoniae or P. aeruginosa. The present sequence is encoded by one of
CC  the target prokaryotic essential genes. Note: The sequence data for this
CC  patent did not form part of the printed specification, but was obtained
CC  in electronic format directly from WIPO at
XX  ftp.wipo.int/pub/published_ptc_sequences
XX
XX  Sequence 299 AA;
SQ
XX
XX  Query Match          5.4%; Score 218.5; DB 6; Length 299;
XX  Best Local Similarity 29.0%; Pred. No. 1.1e-08;
XX  Matches 62; Conservative 29; Mismatches 96; Indels 27; Gaps 6;
XX
XX  551 HRRGHFASVWQSTL-----YVNGELKAPQPMWPKETGYELVSLERWKL 596
XX  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX  17 HSRGKVRHGFPRQSDHSTFTAPLNGFYLLSPALPAQYLPSPR--FPEL-KLLRBEWKT 73
XX
XX  557 TRDESLAMD---KAKLFLPDEENLRKQDWQSFITLMQGRRENNAKQAPKTCUTLLE 652
XX  ||||| : : : : : : : : : : : : : : : : : : : : : :
XX  557 TRDESLAMD---KAKLFLPDEENLRKQDWQSFITLMQGRRENNAKQAPKTCUTLLE 652
```

DB 74 IRDEAFALRDASHIAAATAVANDIGNSFRRNGRRPFLKWKGRPHPSAVALCPRTVELLG 133
 QY 653 KPEETGCRGQIKYSIMHPTGTHWPHGTNCRLRMHGLVTPK-EGCKIRCANETRTW 711
 DB 134 RPSVKAMFQQLP-----PGCRLGLHDPYAGALRYHLGLATPHNDGCAIVDGETYSW 188
 QY 712 BEGKYLIPDDSEHEWODASSFRLLFTVDVWHP 745
 DB 189 RDGEDIVFDETYLHYAFNDQEDRIILFCDIERP 222

RESULT 15
 AAU28081
 ID AAU28081 standard; protein; 324 AA.
 XX
 AC AAU28081;
 XX
 DI 18-DEC-2001 (first entry)
 XX
 DE Novel human secretory protein, Seq ID No 250.
 XX
 KM Human; secreted protein; arthritis; Crohn's disease; sepsis; shock;
 KM ischaemia-reperfusion injury; haematopoiesis; cancer; neuropathy;
 KM transgenic animal; Alzheimer's disease; Parkinson's disease; burn;
 KM amyotrophic lateral sclerosis; platelet disorder; thrombocytopenia;
 KM ulcer; osteoporosis; bone degenerative disorder; periodontal disease;
 KM gut protection; lung; liver fibrosis; immune deficiency; infection;
 KM severe combined immunodeficiency; SCID; autoimmune disorder; allergy;
 KM multiple sclerosis; rheumatoid arthritis; diabetes mellitus; asthma;
 KM fertility; analgesic; pain; antigen.
 XX
 OS Homo sapiens.
 XX
 EN WO200166689-A2.
 PD 13-SEP-2001.
 XX
 PF 05-MAR-2001; 2001WC-US004942.
 XX
 PR 07-MAR-2000; 2000US-00519705.
 PR 19-MAY-2000; 2000US-00574454.
 PR 17-JUN-2000; 2000US-00596193.
 PR 14-JUL-2000; 2000US-00616847.
 PR 19-SEP-2000; 2000US-00665363.
 PR 20-OCT-2000; 2000US-00693267.
 XX
 PA (HYSE-) HXSEQ INC.
 XX
 PI Tang YT, Liu C, Asundi V, Xu C, Wehrman T, Ren F, Ma Y, Zhou F;
 PI Zhao QA, Yang Y, Drmanac RT, Zhang J, Chen R, Xue AJ, Wang J;
 XX
 DR WPI; 2001-589934/66.
 DR N-PSDB; AAS44981.
 XX
 PT Novel polypeptides and nucleic acids obtained from cDNA libraries
 PT prepared from various human tissues, for diagnosis and treatment of
 PT cancer, neurological, inflammatory, and autoimmune disorders.
 XX
 PS Example 4; SEQ ID NO 250; 107bp; English.
 XX
 CC The invention relates to novel isolated human secreted polypeptides (I)
 CC and polynucleotides (II). (I) and (II) are useful for treating
 CC inflammatory conditions such as arthritis, nephritis, Crohn's disease,
 CC ischaemia-reperfusion injury, shock, sepsis, immune responses, and is
 CC involved in increasing haematopoiesis, stem cell survival, bone growth
 CC and remodeling. (I), (II) and modulators of (II) are useful for
 CC prophylaxis or treatment of one or more cancers. (II) is also useful for
 CC creating transgenic animals useful for studying the in vivo activities of
 CC the polypeptide as well as for studying modulators of the polypeptides.
 CC (I) induces the proliferation of neural cells and regeneration of nerve
 CC and brain tissue and is useful for the treatment of central and
 CC peripheral nervous system diseases and neuropathies, such as Alzheimer's,
 CC Parkinson's disease, Huntington's disease, and amyotrophic lateral

CC sclerosis. In addition, (I) is involved in chemotactic or chemokinetic
 CC activity, regulation of haematopoiesis and is useful for treating myeloid
 CC or lymphoid cell disorders, platelet disorders such as thrombocytopenia
 CC and for regeneration of bone, cartilage, tendon, ligament and/or nerve
 CC tissue growth, and in tissue repair, healing of burns, incisions, ulcers,
 CC for treating osteoporosis, osteoarthritis, bone degenerative disorders,
 CC or periodontal disease. Furthermore, (I) is also useful for gut
 CC protection or regeneration and treatment of lung or liver fibrosis,
 CC reperfusion injury in various tissues, various immune deficiencies and
 CC disorders including severe combined immunodeficiency (SCID), bacterial or
 CC fungal infections, autoimmune disorders e.g. multiple sclerosis,
 CC rheumatoid arthritis, diabetes mellitus, myasthenia gravis, allergic
 CC reactions and conditions, such as asthma or other respiratory problems.
 CC In addition, (I) affects biorythms or circadian cycles of rhythms,
 CC fertility, metabolism, catabolism, anabolism, storage or elimination of
 CC dietary fat, lipid, protein, carbohydrate, vitamins, minerals, provides
 CC analgesic effects or other pain reducing effects, immunoglobulin like
 CC activity and can act as an antigen in a vaccine composition to raise an
 CC immune response. AAU28020-AAU28395 represent novel human secreted protein
 CC amino acid sequences of the invention

XX Sequence 324 AA;

Query Match 5.4%; Score 217.5; DB 4; Length 324;
 Best Local Similarity 36.6%; Pred. No. 1.5e-08;
 Matches 53; Conservative 19; Mismatches 56; Indels 17; Gaps 5;

QY 625 WS-----QFTWQGR--RNNACKGAPKTCITLLEKPEPTGCRGQIKYSIMHPG 673
 DB 175 WSPPLAGCYQLLYQAGRCQPSNCRCPGAYRALRGIRSFMSAN--TFGNAGFSVLLDG 232
 QY 674 THWPHGTPTNCRRLRMHGLVTPKESGKIRCANETRTWEGKYLIPDDSEHEWODAS- 732
 DB 233 ARLBGRGCPNNARVRYCHGLKIP-PGCELTVVGGEPQCMABGHCLLVDSFLHTVAHNSP 291
 QY 733 --SFRLLFTVDVWHPBELTPQQRSL 755
 DB 292 EDGPRVVFITVDIMHPNVAGARQAL 316

Search completed: May 19, 2004, 15:57:47
 Job time : 99.9924 secs


```
DB 61 VPPPEAEPTHEPEQVPEAEFPONIEDEAKEQIOSLHEMVAHEVEGEDLQOEDGPJGE 120
QY 179 PQOEDDEFMATVDVDRFETLEPEVSHETESHVEETVSQDCNQDEEMMSQOENPDS 238
DB 121 PQOEDDEFMATVDVDRFETLEPEVSHETESHVEETVSQDCNQDEEMMSQOENPDS 180
QY 239 SEPVEDERLHHDVDDVYQVVEBOAVYEPLENEGIEITETVAPEDNPVEDSQIVAEV 298
DB 181 SEPVEDERLHHDVDDVYQVVEBOAVYEPLENEGIEITETVAPEDNPVEDSQIVAEV 240
QY 299 SIFPVEEQOEVPPEPT 313
DB 241 SIFPVEEQOEVPPEPT 255
```

RESULT 2

```
US-09-702-705-1806
; Sequence 1806, Application US/09702705
; Patent No. 6504010
; GENERAL INFORMATION:
; APPLICANT: Bangur, Tongtong
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Lodes, Michael A.
; APPLICANT: Fanger, Gary
; APPLICANT: Vedvick, Tom
; APPLICANT: Carter, Darlick
; APPLICANT: Retter, Marc
; APPLICANT: Mannion, Jane
; APPLICANT: Fan, Liqun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.478C14
; CURRENT APPLICATION NUMBER: US/09/702,705
; CURRENT FILING DATE: 2000-10-30
; NUMBER OF SEQ ID NOS: 1833
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1806
; LENGTH: 255
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-702-705-1806
```

```
Query Match 32.8%; Score 1320; DB 4; Length 255;
Best Local Similarity 98.8%; Pred. No. 2,1e-99;
Matches 252; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
```

```
QY 59 MWIALLGWTSVAVWPFVLYEVLGKGIYDADGDDPVDVDAKVLLGKERSTSEPA 118
DB 1 MWIALLGWTSVAVWPFVLYEVLGKGIYDADGDDPVDVDAKVLLGKERSTSEPA 60
QY 119 VPPPEAEPTHEPEQVPEAEFPONIEDEAKEQIOSLHEMVAHEVEGEDLQOEDGPJGE 178
DB 61 VPPPEAEPTHEPEQVPEAEFPONIEDEAKEQIOSLHEMVAHEVEGEDLQOEDGPJGE 120
QY 179 PQOEDDEFMATVDVDRFETLEPEVSHETESHVEETVSQDCNQDEEMMSQOENPDS 238
DB 121 PQOEDDEFMATVDVDRFETLEPEVSHETESHVEETVSQDCNQDEEMMSQOENPDS 180
QY 239 SEPVEDERLHHDVDDVYQVVEBOAVYEPLENEGIEITETVAPEDNPVEDSQIVAEV 298
DB 181 SEPVEDERLHHDVDDVYQVVEBOAVYEPLENEGIEITETVAPEDNPVEDSQIVAEV 240
QY 299 SIFPVEEQOEVPPEPT 313
DB 241 SIFPVEEQOEVPPEPT 255
```

```
RESULT 3
US-09-736-457-1806
; Sequence 1806, Application US/09736457
; Patent No. 6509448
; GENERAL INFORMATION:
```

```
APPLICANT: Bangur, Tongtong
APPLICANT: Bangur, Chaitanya S.
APPLICANT: Lodes, Michael A.
APPLICANT: Fanger, Gary
APPLICANT: Vedvick, Tom
APPLICANT: Carter, Darlick
APPLICANT: Retter, Marc
APPLICANT: Mannion, Jane
APPLICANT: Fan, Liqun
APPLICANT: Wang, Aijun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.478C15
; CURRENT APPLICATION NUMBER: US/09/736,457
; CURRENT FILING DATE: 2000-12-13
; NUMBER OF SEQ ID NOS: 1864
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1806
; LENGTH: 255
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-736-457-1806
```

```
Query Match 32.8%; Score 1320; DB 4; Length 255;
Best Local Similarity 98.8%; Pred. No. 2,1e-99;
Matches 252; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
```

```
QY 59 MWIALLGWTSVAVWPFVLYEVLGKGIYDADGDDPVDVDAKVLLGKERSTSEPA 118
DB 1 MWIALLGWTSVAVWPFVLYEVLGKGIYDADGDDPVDVDAKVLLGKERSTSEPA 60
QY 119 VPPPEAEPTHEPEQVPEAEFPONIEDEAKEQIOSLHEMVAHEVEGEDLQOEDGPJGE 178
DB 61 VPPPEAEPTHEPEQVPEAEFPONIEDEAKEQIOSLHEMVAHEVEGEDLQOEDGPJGE 120
QY 179 PQOEDDEFMATVDVDRFETLEPEVSHETESHVEETVSQDCNQDEEMMSQOENPDS 238
DB 121 PQOEDDEFMATVDVDRFETLEPEVSHETESHVEETVSQDCNQDEEMMSQOENPDS 180
QY 239 SEPVEDERLHHDVDDVYQVVEBOAVYEPLENEGIEITETVAPEDNPVEDSQIVAEV 298
DB 181 SEPVEDERLHHDVDDVYQVVEBOAVYEPLENEGIEITETVAPEDNPVEDSQIVAEV 240
QY 299 SIFPVEEQOEVPPEPT 313
DB 241 SIFPVEEQOEVPPEPT 255
```

```
RESULT 4
US-09-671-325-1806
; Sequence 1806, Application US/09671325
; Patent No. 6667154
; GENERAL INFORMATION:
```

```
APPLICANT: Bangur, Tongtong
APPLICANT: Bangur, Chaitanya S.
APPLICANT: Lodes, Michael A.
APPLICANT: Fanger, Gary
APPLICANT: Vedvick, Tom
APPLICANT: Carter, Darlick
APPLICANT: Retter, Marc
APPLICANT: Mannion, Jane
APPLICANT: Fan, Liqun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.478C12
; CURRENT APPLICATION NUMBER: US/09/671,325
; CURRENT FILING DATE: 2000-09-26
; NUMBER OF SEQ ID NOS: 1825
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1806
; LENGTH: 255
; TYPE: PRT
; ORGANISM: Homo sapiens
```

US-09-671-325-1806

Query Match 32.8%; Score 1320; DB 4; Length 255;
Best Local Similarity 98.8%; Pred. No. 2,1e-99;
Matches 252; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 59 MVALLLGWSVAWVDPDLDVYEEVLGKLGTYDADGGDFVYDAAKVLGKERSSTSEPA 118
DB 1 MVALLLGWSVAWVDPDLDVYEEVLGKLGTYDADGGDFVYDAAKVLGKERSSTSEPA 60
QY 119 VPPEAPHTPEEQVPEAEPPQNIIEDEAKQIQLHENVHAEHVGEDLQOEDGPTGE 178
DB 61 VPPEAPHTPEEQVPEAEPPQNIIEDEAKQIQLHENVHAEHVGEDLQOEDGPTGE 120
QY 179 PQGEDFLMATDVDDREFETLEPVSHETSHYVEETVSQDCNQDMEMMSQENPDS 238
DB 121 PQGEDFLMATDVDDREFETLEPVSHETSHYVEETVSQDCNQDMEMMSQENPDS 180
QY 239 SEPVEDEBRLHNDVDTYQYVEEQAVYEPLENGEITETVAPEDNPVEDSQVIVEEV 298
DB 181 SEPVEDEBRLHNDVDTYQYVEEQAVYEPLENGEITETVAPEDNPVEDSQVIVEEV 240
QY 299 SIFPEEQOEVPPEPT 313
DB 241 SIFPEEQOEVPPEPT 255

RESULT 5
5231168-2

Patent No. 5231168
; APPLICANT: DZIEGIEL, MORTEN, BORRE, MARTIN, JESSEN, SOREN;
; VUUST, JENS, RENECK, KLAUS, WIND, ANNETTE, JAKOBSEN, PALE H.
; TITLE OF INVENTION: MALARIA ANTIGEN
; NUMBER OF SEQUENCES: 19
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/409,658
; FILING DATE: 18-SEP-1999
; SEQ ID NO:2:
; LENGTH: 783
5231168-2

Query Match 4.9%; Score 199; DB 6; Length 783;
Best Local Similarity 20.6%; Pred. No. 1,6e-07;
Matches 143; Conservative 134; Mismatches 237; Indels 180; Gaps 36;

QY 81 EEVLGKGIYDADGGDF-----DVDDAKVL---GLKERSSTSEPAVPE--EAE 125
DB 119 DEESGEGVLDNE--EGDFEENHEEFPPDQNDSELSNELVESKVSSEPAHEVIVSEK 177
QY 126 PHTPEEQVPEAEPPQNIIEDEAKQIQLHENVHAEHVGEDLQOEDGPTG----- 177
DB 178 SVSEPAHEVIVSEKST--SEPAHEVSV-----SQSNNEPEKKDGPVPSKPEIE 229
QY 178 ----EPOEDDEFMATDVDDREFETLEP-----EVSHETSHYVEETVSQDCNQ 224
DB 230 KVDVQPKIVDLQITEPNFVDSQPNPQBPVESEVYKIEKVSEENKHA--SVDPVEKE--KE 286
QY 225 DMEHMEQENPDSSE---PVVEDERLHNDVDTYQYVEEQAVYEPLENGEITETV 280
DB 287 NVSVEVEKQNSQSVSEIVPNEDE-----FEDVHTQDLDDHKTVP---EIVEVEIIP 338
QY 281 APEPDNVEDSQVT--VEEVSIFP-----VEEQOEVPPEPTNRK 316
DB 339 SELHENEVAHPEIVEIEEV--FPEPNQNNFEQIENEDKSAHQHIVEVEIIPBEDKN 396
QY 317 TDDEQAKAYKKKKKXKLKFKDKIKAEIDAAE-----KLRKKGIEMVANNKELVYRK 370
DB 397 EKVEHETVEVEEILPE-----DKKEKGQHEIVEVEIIPBEDKNKKEHETVEVEIIP 451
QY 371 YPQSPRAYGKACQEDDLAEKRSNEVLRGAIETTYQEVASLPDVPAD-----LKLK 422
DB 452 -DKNEKGQHEIVEVEIIPEDK--NEKVEHETVEVEI-----LPEDKNKGQHEIVEVE 503

QY 423 LKRRSPQOFLGMKRGSLTLQRLVOLFPNDTSLKNDLG-----VGYLLIGDNDNAKV 476
DB 504 EILPEDKNEKVOH-----EIVEVEIIPEDK-----KNEKGQHEIVEVEIIPEDKNEKQ 554
QY 477 YE--EVLASVPNGFAKVNHGFI-----LKAQNIAISIPYKXGIESGDGDT 523
DB 555 HEIVEVEIIPEDKNEKVOHEIVEVEIIPEDKNEKQHEIVEVEIIPEDKNEKQHEIVEVEIIPEDK 607
QY 524 DGRFYEFLGAMORVGNKAEKAYVELGKRGHFAVQBSLYN--VNGLKAQPMWTPK- 579
DB 608 -----IEEVPQNTNNENIETIKPEKNOEF--SVEEKAIQBPVPTLNENENVTPK 659
QY 580 --ETGTEILVKSLENNKILRD----EGLAVMDKAKGL--EIPEDENTREKGDMSQFT 629
DB 660 SESESTPDIYQI-----KIVQENKPKKETTPTVDGPKHVEQNIQEDDNDDEDDDDIDFE 714
QY 630 LW-----QOGRRENKAGAPKTCITLLEKPEPT 658
DB 715 GLSRKDEKDSNNKKKSSFTYISTKRRKVS 748

RESULT 6

US-09-252-991A-29355
; Sequence 29355; Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196,136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 29355
; LENGTH: 320
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-29355

Query Match 4.9%; Score 197.5; DB 4; Length 320;
Best Local Similarity 25.9%; Pred. No. 5,7e-08;
Matches 58; Conservative 34; Mismatches 85; Indels 47; Gaps 9;

QY 551 HKRGHFAVQBSLYNNGLKAQPMWTPKETGYTEL-----VKSL 590
DB 31 HFRGRARLPPLRLQLVNHA-----WPAF--YNSLWYLFSSVPSKLYLDRSPPELD 81
QY 591 ERWVKILRDGLVMDKAKGLFPEDENTREKQ-----DWSQFTL--WQGRRENKACX 642
DB 82 KNNWQITREBALNLPBEG--YIRALNNNEAGFSFFKGWKRFYLTWDGFL--PSAQ 137
QY 643 GAPRTCTLEKPEPTTGCRRGQIKYSIMAPGTVMPHTGPNCRRLNMLGLVLP--KSGCK 701
DB 138 LCKRTVELVSRINVGGA-----MFTLLPFGSHLPNPRDPFGSLRYHLGLSTNSNCR 192
QY 702 IRCANETRWBEKGLVIFDPSFEHVAQDASSRLFTVDVWHP 745
DB 193 IYVDGQYAWRDSGDVWFDETFVHWVKNTEQTRVILFCIDIERP 236

RESULT 7

US-08-072-610-2
; Sequence 2; Application US/08072610
; Patent No. 5532133
; GENERAL INFORMATION:
; APPLICANT: Barnwell, John
; TITLE OF INVENTION: Plasmidium vivax Blood Stage Antigens,
; TITLE OF INVENTION: Monoclonal Antibodies, and Diagnostic Assays
; NUMBER OF SEQUENCES: 2

QY 356 KIEBANAFAKELVRYKYPQ---SEPAAYGAQCCDDIAEKRRSNEVLGALITYOEAVSL 411
Db 913 ELIEVEEQOGVYQCEQETEBEILIHGSSS-EDEN-----EVDYVVSHTHEQIASS 964
QY 412 PDVAD 417
Db 965 P--PGD 968

RESULT 13

US-09-328-352-8015
Sequence 8015, Application US/09328352
Patent No. 6562958
GENERAL INFORMATION:
APPLICANT: Gary L. Brelton et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
FILE REFERENCE: GTC99-03PA
CURRENT APPLICATION NUMBER: US/09/328,352
CURRENT FILING DATE: 1999-06-04
NUMBER OF SEQ ID NOS: 8252
SEQ ID NO 8015
LENGTH: 312
TYPE: PR1
ORGANISM: Acinetobacter baumannii
US-09-328-352-8015

Query Match 4.7%; Score 189.5; DB 4; Length 312;
Best Local Similarity 27.8%; Pred. No. 2.5e-07;
Matches 52; Conservative 34; Mismatches 88; Indels 13; Gaps 4;

QY 564 LYNVNGLKQAPWMTPKETGYTELKSLERNMKLIRDEGLAVMD---KAKGLFEPEDENL 619
Db 53 MYIFSVPNPY---IDTQHFQDLKYLDEWEMIRBAKALYQGGKASSTYNDIGFMS 109
QY 620 REKGWDSQFTLWQGGRRNENACKGAPYCTTLEKFEFTTGCRRGQIKYSIMPGTHWDP 679
Db 110 FFKTGKRRFYIKMYESSHSAELCEKTXALTKLTPTIKAA-----MFTELAPDSRLVNH 164
QY 680 TGTNCRRLRHLGLVLPKESG-CIKRCANETRTWECKVLIFDPSFEHFWQDASSRLLF 738
Db 165 RDPYASGLRHLHLIFLINDRCFLDYDGERYSWRDGSVDFETYLHVANKTDQNRILF 224

QY 739 IVDVWHP 745
Db 225 FADVERP 231

RESULT 14

US-08-961-083-218
Sequence 218, Application US/08961083
Patent No. 6159469
GENERAL INFORMATION:
APPLICANT: Choi et. al.
TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
NUMBER OF SEQUENCES: 452
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/961,083
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:

APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 218:
SEQUENCE CHARACTERISTICS:
LENGTH: 565 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-961-083-218

Query Match 4.4%; Score 177.5; DB 3; Length 565;
Best Local Similarity 22.5%; Pred. No. 5.5e-06;
Matches 114; Conservative 71; Mismatches 199; Indels 123; Gaps 22;

QY 111 ERSTSEPAVPPEE--AEPTPEEQVPEAEQNIEDAKKQIQSLHENVNAHVEGED 168
Db 62 KEDSABPA-PVEEVGGEVSEKPEKVAVKPEQSPDKPABESKVPQAGEPV---APRED 116
QY 169 LQEDSGTEBPOQDEDFLMATDVDRFETLEPEVSHETSHYHVEEYVSQDQNC-DWE 227
Db 117 EKAPVEBKKPEKPEE--KAVEETPKQESTPDTAEBETVED--KEETVNOISTEOPKVE 172
QY 228 EMMSQENPDSSSEPVVEDE---RLHHDVDVYQYEEQAVPEPLENKGEITEVTAP 283
Db 173 TPAVEKQTEPTPEPKVEQGEVAPAREDEQAPTAPVPEPKQEPVPEEKAV---EETPK 229
QY 284 ED-----NPEDSQV---IVEYSIFPV-----EQQ 307
Db 230 EDKIKIGTKEVDSSELNNQIDKASVSAPTSTASYNALGVLLETAKGVASPEVKOP 289
QY 308 EYVPEETNR-----KTD-----DPEKAKVKKKKPKLNFEDTIRKELDAEK 350
Db 290 EVNSETNKKTLIDALNVDTKELNNTIADAKTKVHSHYSDRSMQNTQTEYTAEKYAAANT 349
QY 351 LRRKGIEBANAFA-----XELVRYKYPQSPARAYGKAQ 383
Db 350 DAKQSEVNAVAVEKLTATIEKLVLESEKPIITLSTDKILERRAAVKYLE---NQNKTK 406
QY 384 CEDDLAEKRSNEVLGALITYOEVASLPDVPADILKLSKRRSDRQQLGHRGSLTI 443
Db 407 IKSITAEIKKGEHIVNTVLTDKVT--ETISAFAKNLEYYKXYTLSTMIYDRNGEBT 465
QY 444 QRL-----VQFPNDLSLKY---DLGAGYLLIGDNDNAKKVEVYLVTPNDGFARVHG 495
Db 466 ETLNQNITQDLKVKELKNIKRTD--IKY-----ENKGTWESLITTPD--KSNYY 515
QY 496 FILKQNKIAE--SIPYLKEGIESGDP 520
Db 516 LKITSNNQKTTLLAVNIEETTVNGTP 542

RESULT 15

US-09-536-784-218
Sequence 218, Application US/09536784
Patent No. 6573082
GENERAL INFORMATION:

APPLICANT: Choi et. al.
TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
NUMBER OF SEQUENCES: 452
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 19, 2004, 15:59:55 ; Search time 74.9354 Seconds

(without alignments)
2814.722 Million cell updates/sec

Title: US-09-903-199-2

Sequence: 1 MAQRKNAKSSGNSSSGSGS.....IVDVHPELTPQRRSLPAI 758

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1145568 seqs, 278261457 residues

Total number of hits satisfying chosen parameters: 1145568

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database:

1: Published Applications AA:
2: /cgn2_6/ptodata/2/pubppa/US07_PUBCOMB.pep:
3: /cgn2_6/ptodata/2/pubppa/US06_NEW_PUB.pep:
4: /cgn2_6/ptodata/2/pubppa/US06_PUBCOMB.pep:
5: /cgn2_6/ptodata/2/pubppa/US07_NEW_PUB.pep:
6: /cgn2_6/ptodata/2/pubppa/US08_NEW_PUB.pep:
7: /cgn2_6/ptodata/2/pubppa/US08_PUBCOMB.pep:
8: /cgn2_6/ptodata/2/pubppa/US09_PUBCOMB.pep:
9: /cgn2_6/ptodata/2/pubppa/US09_PUBCOMB.pep:
10: /cgn2_6/ptodata/2/pubppa/US09_PUBCOMB.pep:
11: /cgn2_6/ptodata/2/pubppa/US09_PUBCOMB.pep:
12: /cgn2_6/ptodata/2/pubppa/US10A_PUBCOMB.pep:
13: /cgn2_6/ptodata/2/pubppa/US10A_PUBCOMB.pep:
14: /cgn2_6/ptodata/2/pubppa/US10C_PUBCOMB.pep:
15: /cgn2_6/ptodata/2/pubppa/US10C_PUBCOMB.pep:
16: /cgn2_6/ptodata/2/pubppa/US10_NEW_PUB.pep:
17: /cgn2_6/ptodata/2/pubppa/US60_NEW_PUB.pep:
18: /cgn2_6/ptodata/2/pubppa/US60_PUBCOMB.pep:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4022	100.0	758	US-09-903-248-2	Sequence 2, Appli
2	4022	100.0	758	US-09-859-604-2	Sequence 2, Appli
3	4022	100.0	758	US-09-903-063-2	Sequence 2, Appli
4	4022	100.0	758	US-09-903-216-2	Sequence 2, Appli
5	4022	100.0	758	US-09-903-199-2	Sequence 2, Appli
6	4022	100.0	758	US-09-903-023-2	Sequence 2, Appli
7	4022	100.0	758	US-09-436-184-2	Sequence 2, Appli
8	1320	32.8	255	US-09-736-457-1806	Sequence 1806, Ap
9	1320	32.8	255	US-09-902-941-1806	Sequence 1806, Ap
10	1320	32.8	255	US-09-849-626-1806	Sequence 1806, Ap
11	1320	32.8	255	US-10-283-017-1806	Sequence 1806, Ap
12	1320	32.8	255	US-10-017-754-1806	Sequence 1806, Ap
13	1320	32.8	255	US-10-113-872-1806	Sequence 1806, Ap
14	275	6.8	250	US-10-156-761-8153	Sequence 8153, Ap
15	274.5	6.8	369	US-10-168-274-15	Sequence 15, Appli

16	218.5	5.4	299	12	US-10-282-122A-49023	Sequence 49023, A
17	217.5	5.4	324	12	US-10-221-278-250	Sequence 250, App
18	217.5	5.4	324	15	US-10-291-172-250	Sequence 250, App
19	199	4.9	109	14	US-10-168-274-5	Sequence 5, Appli
20	196.5	4.9	355	12	US-10-282-122A-51120	Sequence 51120, A
21	190.5	4.7	1162	12	US-09-894-273-2	Sequence 2, Appli
22	189.5	4.7	1162	14	US-10-294-804-2	Sequence 2, Appli
23	189.5	4.7	304	12	US-10-282-122A-45188	Sequence 45188, A
24	189.5	4.7	1616	10	US-09-820-843A-16	Sequence 16, Appli
25	189.5	4.7	1616	12	US-10-282-122A-63593	Sequence 63593, A
26	186	4.6	50	16	US-10-443-622-48	Sequence 48, Appli
27	182	4.5	350	13	US-10-101-487-58	Sequence 58, Appli
28	177.5	4.4	565	9	US-09-765-272-218	Sequence 218, App
29	177	4.4	240	13	US-10-101-487-75	Sequence 75, Appli
30	177	4.4	845	12	US-10-205-331-61	Sequence 61, Appli
31	175	4.4	1000	14	US-10-128-714-3105	Sequence 3305, Ap
32	174	4.3	566	12	US-10-424-599-172054	Sequence 172054, A
33	172.5	4.3	1302	12	US-10-282-122A-46112	Sequence 46112, A
34	170	4.2	299	12	US-10-282-122A-50678	Sequence 50678, A
35	168.5	4.2	200	13	US-10-101-487-53	Sequence 53, Appli
36	168.5	4.2	300	12	US-10-282-122A-60256	Sequence 60256, A
37	168	4.2	302	12	US-10-282-122A-75844	Sequence 75844, A
38	167.5	4.2	1876	12	US-10-282-122A-73861	Sequence 73861, A
39	167	4.2	187	13	US-10-101-487-50	Sequence 50, Appli
40	167	4.2	302	9	US-09-815-242-13815	Sequence 13815, A
41	165.5	4.1	1095	14	US-10-128-714-8305	Sequence 8305, Ap
42	165	4.1	518	12	US-10-425-114-71472	Sequence 71472, A
43	164	4.1	1957	15	US-10-369-493-2070	Sequence 2070, Ap
44	163.5	4.1	1297	12	US-10-282-122A-71692	Sequence 71692, A
45	163.5	4.1	1501	9	US-09-924-154-17	Sequence 17, Appli

ALIGNMENTS

RESULT 1
US-09-903-248-2
Sequence 2, Application US/09903248
Patent No. US20020102263A1
GENERAL INFORMATION:
APPLICANT: Wanda, Jack R.
APPLICANT: de la Monte, Suzanne M.
APPLICANT: Ince, Nedim
APPLICANT: Carlson, Rolf I.
TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF MALIGNANT NEOPLASMS
FILE REFERENCE: 21486-032 DIV5
CURRENT APPLICATION NUMBER: US/09/903,248
CURRENT FILING DATE: 2001-07-11
PRIOR APPLICATION NUMBER: 09/436,184
PRIOR FILING DATE: 1999-11-08
NUMBER OF SEQ ID NOS: 9
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2
LENGTH: 758
TYPE: PRT
ORGANISM: Homo sapiens
US-09-903-248-2

Query Match 100.0%, Score 4022, DB 9, Length 758;
Best Local Similarity 100.0%, Pred. No. 1.56-282;
Matches 758; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
1 MAQRKNAKSSGNSSSGSGSGSPPARRRTKGGHNGKKGISGTSFTFWFW 60
1 IALLVMTSVAVVWFDLVYEVLLKGLGYDADGGDPVDVDAKYLGLKERTSPAPV 120
61 IALLVMTSVAVVWFDLVYEVLLKGLGYDADGGDPVDVDAKYLGLKERTSPAPV 120
61 IALLVMTSVAVVWFDLVYEVLLKGLGYDADGGDPVDVDAKYLGLKERTSPAPV 120
121 PEEAPPTPEPEQVVEAPQNTIEDEAKEQIQLIHENYHAHVGEEDLQEDGPTGEQ 180
121 PEEAPPTPEPEQVVEAPQNTIEDEAKEQIQLIHENYHAHVGEEDLQEDGPTGEQ 180
121 PEEAPPTPEPEQVVEAPQNTIEDEAKEQIQLIHENYHAHVGEEDLQEDGPTGEQ 180

QY 181 QEDDEFLMATDVDRFETLEPEVSHETESHVHEETVSQDCNODMEEMMSSEQENPSSSE 240
DB 181 QEDDEFLMATDVDRFETLEPEVSHETESHVHEETVSQDCNODMEEMMSSEQENPSSSE 240
QY 241 PVEDERLHDDTDVTVQVVEBAVVEPLENEGIEITEVTAPREDNVEDVSQVVEVSI 300
DB 241 PVEDERLHDDTDVTVQVVEBAVVEPLENEGIEITEVTAPREDNVEDVSQVVEVSI 300
QY 301 FPEVEQOEVPETNRKTDDEQAKAVKKKKPKLNFCDKTIKALDAEKLRKGIIEEA 360
DB 301 FPEVEQOEVPETNRKTDDEQAKAVKKKKPKLNFCDKTIKALDAEKLRKGIIEEA 360
QY 361 VNAKELVRKYPOSPPRAYGKAQCEDDLAEKRSNEVLRGAIETTYOVASLPVPADLLK 420
DB 361 VNAKELVRKYPOSPPRAYGKAQCEDDLAEKRSNEVLRGAIETTYOVASLPVPADLLK 420
QY 421 LSLKRRSDROQFLGHRGSLTLTQRLVOLFPNDTSLKNDLGVGXLLIGDNDNAKRYEEV 480
DB 421 LSLKRRSDROQFLGHRGSLTLTQRLVOLFPNDTSLKNDLGVGXLLIGDNDNAKRYEEV 480
QY 481 LSVTPNDGFAKVHYGFLKAQNKIAESIPLYKEGIESGDPGTDDGRFYFHLGDAMORVGN 540
DB 481 LSVTPNDGFAKVHYGFLKAQNKIAESIPLYKEGIESGDPGTDDGRFYFHLGDAMORVGN 540
QY 541 KEAKYWEIGHKRGHFAVWQORSLYVNGLKAOPWMTPKETGYTELKSLERWKLIRDE 600
DB 541 KEAKYWEIGHKRGHFAVWQORSLYVNGLKAOPWMTPKETGYTELKSLERWKLIRDE 600
QY 601 GLAVMDKAKGLFLPEBENLRKGDMSQFTLMQGRNENACKAPKCTLLLEKPEPTGC 660
DB 601 GLAVMDKAKGLFLPEBENLRKGDMSQFTLMQGRNENACKAPKCTLLLEKPEPTGC 660
QY 661 RRGQIKYSIMHPGTHWPHPTGPTNCRRLRMHLGLVTPREGCKIRCANETRWEGKVLIFD 720
DB 661 RRGQIKYSIMHPGTHWPHPTGPTNCRRLRMHLGLVTPREGCKIRCANETRWEGKVLIFD 720
QY 721 DSFEHEVWODASSFRLLIFIVDWMHPELTPQORSLPAI 758
DB 721 DSFEHEVWODASSFRLLIFIVDWMHPELTPQORSLPAI 758

RESULT 2
US-09-859-604-2
; Sequence 2, Application US/09859604
; Patent No. US20020110559A1
; GENERAL INFORMATION:
; APPLICANT: Wands, Jack R.
; APPLICANT: de la Monte, Suzanne M
; APPLICANT: Deubart, Hoessein A
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF MALIGNANT NEOPLASMS
; FILE REFERENCE: 21486-032 CIP
; CURRENT APPLICATION NUMBER: US/09/859,604
; CURRENT FILING DATE: 2001-05-17
; PRIOR APPLICATION NUMBER: 09/436,184
; PRIOR FILING DATE: 1999-11-08
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 758
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-859-604-2

Query Match 100.0%; Score 4022; DB 9; Length 758;
Best Local Similarity 100.0%; Pred. No. 1,5e-282;
Matches 758; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAORKKAKSSGSSSSGSGSAGSSPGARRTHGCHKNRKGLSGTSFTWPMV 60
DB 1 MAORKKAKSSGSSSSGSGSAGSSPGARRTHGCHKNRKGLSGTSFTWPMV 60

QY 61 IALLGWTSVAVWMEPLVNDYEEVLGKLGITYDADGDGDFVDAKALLGLKERSTSEPAV 120
DB 61 IALLGWTSVAVWMEPLVNDYEEVLGKLGITYDADGDGDFVDAKALLGLKERSTSEPAV 120
QY 121 PEEAEPHTEPEEOPVPAEAPONIEDAKEQIOSLHEMVAABVEGEDLOQEDGPTGEPQ 180
DB 121 PEEAEPHTEPEEOPVPAEAPONIEDAKEQIOSLHEMVAABVEGEDLOQEDGPTGEPQ 180
QY 181 QEDDEFLMATDVDRFETLEPEVSHETESHVHEETVSQDCNODMEEMMSSEQENPSSSE 240
DB 181 QEDDEFLMATDVDRFETLEPEVSHETESHVHEETVSQDCNODMEEMMSSEQENPSSSE 240
QY 241 PVEDERLHDDTDVTVQVVEBAVVEPLENEGIEITEVTAPREDNVEDVSQVVEVSI 300
DB 241 PVEDERLHDDTDVTVQVVEBAVVEPLENEGIEITEVTAPREDNVEDVSQVVEVSI 300
QY 301 FPEVEQOEVPETNRKTDDEQAKAVKKKKPKLNFCDKTIKALDAEKLRKGIIEEA 360
DB 301 FPEVEQOEVPETNRKTDDEQAKAVKKKKPKLNFCDKTIKALDAEKLRKGIIEEA 360
QY 361 VNAKELVRKYPOSPPRAYGKAQCEDDLAEKRSNEVLRGAIETTYOVASLPVPADLLK 420
DB 361 VNAKELVRKYPOSPPRAYGKAQCEDDLAEKRSNEVLRGAIETTYOVASLPVPADLLK 420
QY 421 LSLKRRSDROQFLGHRGSLTLTQRLVOLFPNDTSLKNDLGVGXLLIGDNDNAKRYEEV 480
DB 421 LSLKRRSDROQFLGHRGSLTLTQRLVOLFPNDTSLKNDLGVGXLLIGDNDNAKRYEEV 480
QY 481 LSVTPNDGFAKVHYGFLKAQNKIAESIPLYKEGIESGDPGTDDGRFYFHLGDAMORVGN 540
DB 481 LSVTPNDGFAKVHYGFLKAQNKIAESIPLYKEGIESGDPGTDDGRFYFHLGDAMORVGN 540
QY 541 KEAKYWEIGHKRGHFAVWQORSLYVNGLKAOPWMTPKETGYTELKSLERWKLIRDE 600
DB 541 KEAKYWEIGHKRGHFAVWQORSLYVNGLKAOPWMTPKETGYTELKSLERWKLIRDE 600
QY 601 GLAVMDKAKGLFLPEBENLRKGDMSQFTLMQGRNENACKAPKCTLLLEKPEPTGC 660
DB 601 GLAVMDKAKGLFLPEBENLRKGDMSQFTLMQGRNENACKAPKCTLLLEKPEPTGC 660
QY 661 RRGQIKYSIMHPGTHWPHPTGPTNCRRLRMHLGLVTPREGCKIRCANETRWEGKVLIFD 720
DB 661 RRGQIKYSIMHPGTHWPHPTGPTNCRRLRMHLGLVTPREGCKIRCANETRWEGKVLIFD 720
QY 721 DSFEHEVWODASSFRLLIFIVDWMHPELTPQORSLPAI 758
DB 721 DSFEHEVWODASSFRLLIFIVDWMHPELTPQORSLPAI 758

RESULT 3
US-09-903-063-2
; Sequence 2, Application US/09903063
; Patent No. US20020114810A1
; GENERAL INFORMATION:
; APPLICANT: Wands, Jack R.
; APPLICANT: de la Monte, Suzanne M.
; APPLICANT: Ince, Nedim
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF MALIGNANT NEOPLASMS
; FILE REFERENCE: 21486-032 DIV3
; CURRENT APPLICATION NUMBER: US/09/903,063
; CURRENT FILING DATE: 2001-10-11
; PRIOR APPLICATION NUMBER: 09/436,184
; PRIOR FILING DATE: 1999-11-08
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 758
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-903-063-2

Query Match 100.0%; Score 4022; DB 9; Length 758;

Best Local Similarity 100.0%; Pred. No. 1.5e-282;
Matches 758; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 MAQRKAKSSGSSSSGSGSTAGSSSPGARRETYHGHNKRKGGSGTSFTFMV 60
Db 1 MAQRKAKSSGSSSSGSGSTAGSSSPGARRETYHGHNKRKGGSGTSFTFMV 60
QY 61 IALLGWTSAVAVWFDLVYEEVLGKLGIDADGDGDFVDADAVLLGLKERSTSEPAVP 120
Db 61 IALLGWTSAVAVWFDLVYEEVLGKLGIDADGDGDFVDADAVLLGLKERSTSEPAVP 120
QY 121 PEEAEPHTPEBEQVPVEAEPOQIDEAKEOIOSLHEMVAHEVGEEDLQOEDGPTGEPQ 180
Db 121 PEEAEPHTPEBEQVPVEAEPOQIDEAKEOIOSLHEMVAHEVGEEDLQOEDGPTGEPQ 180
QY 181 QEDDEFMATDVDRFETLEPEVSHETESHVYEETVSODCNQDMEMMSSEQENPDSE 240
Db 181 QEDDEFMATDVDRFETLEPEVSHETESHVYEETVSODCNQDMEMMSSEQENPDSE 240
QY 241 PVEDEBLHHDVDVTYQVVEEQAVVEPLENEGIEITEVTAPPEDNVEDSOVIVEEVS1 300
Db 241 PVEDEBLHHDVDVTYQVVEEQAVVEPLENEGIEITEVTAPPEDNVEDSOVIVEEVS1 300
QY 301 FVVEEQOEVPETNRKTDDEQAKAYKKKPKLINFEDKTIKALDAEKLRRKGLIEA 360
Db 301 FVVEEQOEVPETNRKTDDEQAKAYKKKPKLINFEDKTIKALDAEKLRRKGLIEA 360
QY 361 VNAFKELVRYKYPSPARAYGKAQCEDDLAEKRSNEVLRAIETTYGEVASLDPVPADLTK 420
Db 361 VNAFKELVRYKYPSPARAYGKAQCEDDLAEKRSNEVLRAIETTYGEVASLDPVPADLTK 420
QY 421 LSLKRSDRQOFLGHNMGSLTLQRLVOLFPNDTSLKNDIGVGYLLIGDNDNAKYYEEV 480
Db 421 LSLKRSDRQOFLGHNMGSLTLQRLVOLFPNDTSLKNDIGVGYLLIGDNDNAKYYEEV 480
QY 481 LSVTPNDGFAKVHYGFLTKAONKIASIPLYKGISGDPGTDDGRFYFHLGDMORVGN 540
Db 481 LSVTPNDGFAKVHYGFLTKAONKIASIPLYKGISGDPGTDDGRFYFHLGDMORVGN 540
QY 541 KEAYKWELEGHKGHGFASVWQRSLYNVNGIKAQPMWTPKETGYTELKVSLEKRWKLI RDE 600
Db 541 KEAYKWELEGHKGHGFASVWQRSLYNVNGIKAQPMWTPKETGYTELKVSLEKRWKLI RDE 600
QY 601 GLAVMDKAGLFLPEDENLREKGDMSOFTLMQOGRNENNA CKAPKTCITLLEKFPETTG 660
Db 601 GLAVMDKAGLFLPEDENLREKGDMSOFTLMQOGRNENNA CKAPKTCITLLEKFPETTG 660
QY 661 RRQIKYSIMHPGTHWPHGPTNCRRLRMHGLVIRPREGCKIRCANETRTWEGSKVLI RDE 720
Db 661 RRQIKYSIMHPGTHWPHGPTNCRRLRMHGLVIRPREGCKIRCANETRTWEGSKVLI RDE 720
QY 721 DSFEHEWMDASSFRLLIFIVDVWHPBELTPOQRSLPAI 758
Db 721 DSFEHEWMDASSFRLLIFIVDVWHPBELTPOQRSLPAI 758

```

RESULT 4
US-09-903-216-2

```

; Sequence 2, Application US/09903216
; Patent No. US20020114811A1
; GENERAL INFORMATION:
; APPLICANT: Mands, Jack R.
; APPLICANT: de la Monte, Suzanne M.
; APPLICANT: Ince, Nedim
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF MALIGNANT NEOPLASMS
; FILE REFERENCE: 21486-032 DIV2
; CURRENT APPLICATION NUMBER: US/09/903,216
; PRIOR FILING DATE: 2001-07-11
; PRIOR APPLICATION NUMBER: 09/436,184
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn Ver. 2.1

```

; SEQ ID NO 2
; LENGTH: 758
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-903-216-2

Query Match 100.0%; Score 4022; DB 9; Length 758;
Best Local Similarity 100.0%; Pred. No. 1.5e-282;
Matches 758; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 MAQRKAKSSGSSSSGSGSTAGSSSPGARRETYHGHNKRKGGSGTSFTFMV 60
Db 1 MAQRKAKSSGSSSSGSGSTAGSSSPGARRETYHGHNKRKGGSGTSFTFMV 60
QY 61 IALLGWTSAVAVWFDLVYEEVLGKLGIDADGDGDFVDADAVLLGLKERSTSEPAVP 120
Db 61 IALLGWTSAVAVWFDLVYEEVLGKLGIDADGDGDFVDADAVLLGLKERSTSEPAVP 120
QY 121 PEEAEPHTPEBEQVPVEAEPOQIDEAKEOIOSLHEMVAHEVGEEDLQOEDGPTGEPQ 180
Db 121 PEEAEPHTPEBEQVPVEAEPOQIDEAKEOIOSLHEMVAHEVGEEDLQOEDGPTGEPQ 180
QY 181 QEDDEFMATDVDRFETLEPEVSHETESHVYEETVSODCNQDMEMMSSEQENPDSE 240
Db 181 QEDDEFMATDVDRFETLEPEVSHETESHVYEETVSODCNQDMEMMSSEQENPDSE 240
QY 241 PVEDEBLHHDVDVTYQVVEEQAVVEPLENEGIEITEVTAPPEDNVEDSOVIVEEVS1 300
Db 241 PVEDEBLHHDVDVTYQVVEEQAVVEPLENEGIEITEVTAPPEDNVEDSOVIVEEVS1 300
QY 301 FVVEEQOEVPETNRKTDDEQAKAYKKKPKLINFEDKTIKALDAEKLRRKGLIEA 360
Db 301 FVVEEQOEVPETNRKTDDEQAKAYKKKPKLINFEDKTIKALDAEKLRRKGLIEA 360
QY 361 VNAFKELVRYKYPSPARAYGKAQCEDDLAEKRSNEVLRAIETTYGEVASLDPVPADLTK 420
Db 361 VNAFKELVRYKYPSPARAYGKAQCEDDLAEKRSNEVLRAIETTYGEVASLDPVPADLTK 420
QY 421 LSLKRSDRQOFLGHNMGSLTLQRLVOLFPNDTSLKNDIGVGYLLIGDNDNAKYYEEV 480
Db 421 LSLKRSDRQOFLGHNMGSLTLQRLVOLFPNDTSLKNDIGVGYLLIGDNDNAKYYEEV 480
QY 481 LSVTPNDGFAKVHYGFLTKAONKIASIPLYKGISGDPGTDDGRFYFHLGDMORVGN 540
Db 481 LSVTPNDGFAKVHYGFLTKAONKIASIPLYKGISGDPGTDDGRFYFHLGDMORVGN 540
QY 541 KEAYKWELEGHKGHGFASVWQRSLYNVNGIKAQPMWTPKETGYTELKVSLEKRWKLI RDE 600
Db 541 KEAYKWELEGHKGHGFASVWQRSLYNVNGIKAQPMWTPKETGYTELKVSLEKRWKLI RDE 600
QY 601 GLAVMDKAGLFLPEDENLREKGDMSOFTLMQOGRNENNA CKAPKTCITLLEKFPETTG 660
Db 601 GLAVMDKAGLFLPEDENLREKGDMSOFTLMQOGRNENNA CKAPKTCITLLEKFPETTG 660
QY 661 RRQIKYSIMHPGTHWPHGPTNCRRLRMHGLVIRPREGCKIRCANETRTWEGSKVLI RDE 720
Db 661 RRQIKYSIMHPGTHWPHGPTNCRRLRMHGLVIRPREGCKIRCANETRTWEGSKVLI RDE 720
QY 721 DSFEHEWMDASSFRLLIFIVDVWHPBELTPOQRSLPAI 758
Db 721 DSFEHEWMDASSFRLLIFIVDVWHPBELTPOQRSLPAI 758

```

RESULT 5
US-09-903-199-2

```

; Sequence 2, Application US/09903199
; Patent No. US20020122802A1
; GENERAL INFORMATION:
; APPLICANT: Mands, Jack R.
; APPLICANT: de la Monte, Suzanne M.
; APPLICANT: Ince, Nedim
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF MALIGNANT NEOPLASMS

```


Db 721 DSFEHEWODASSFRLIFFIVDWMHPELTPOQRSLPAI 758

RESULT 7

US-09-436-184-2
 ; Sequence 2, Application US/09436184
 ; Publication No. US20030031670A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Wands, Jack R.
 ; APPLICANT: de la Monte, Suzanne M.
 ; APPLICANT: Ince, Nedim
 ; APPLICANT: Carlson, Rolf I.
 ; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF MALIGNANT NEOPLASMS
 ; FILE REFERENCE: R.I. Hosp. - Malignant Neoplasms
 ; CURRENT APPLICATION NUMBER: US/09/436,184
 ; CURRENT FILING DATE: 1999-11-08
 ; NUMBER OF SEQ ID NOS: 7
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 2
 ; LENGTH: 758
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-09-436-184-2

Query Match 100.0%; Score 4022; DB 10; Length 758;
 Best Local Similarity 100.0%; Pred. No. 1.5e-282;
 Matches 758; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAQRKNAKSSGNSSSSGSGSGSTAGSSSPGARRETKHGKHNGKRGKGLSGTSFTFWNV 60
 DB 1 MAQRKNAKSSGNSSSSGSGSGSTAGSSSPGARRETKHGKHNGKRGKGLSGTSFTFWNV 60
 QY 61 IALIGWTSVAVVMFPLVJVEEYLGKLGIVDADGDFVDADAKVLLGLKERSTSEPAVP 120
 DB 61 IALIGWTSVAVVMFPLVJVEEYLGKLGIVDADGDFVDADAKVLLGLKERSTSEPAVP 120
 QY 121 PEBAEPHTEPEEVPVPAEAPONTEDBAKQIOSLHEMVAHEVEGEDLOQEDGPTEBPQ 180
 DB 121 PEBAEPHTEPEEVPVPAEAPONTEDBAKQIOSLHEMVAHEVEGEDLOQEDGPTEBPQ 180
 QY 181 QEDDEFIATDVDDRFETLEPEVSHETESHVHEETVSQDCNQDMEMNSEQENDPSS 240
 DB 181 QEDDEFIATDVDDRFETLEPEVSHETESHVHEETVSQDCNQDMEMNSEQENDPSS 240
 QY 241 PVVEDERLHHDNDVTVYVEEQAAYVEPLENEGIEITETVAPREDNPVEDSQVIVEEVI 300
 DB 241 PVVEDERLHHDNDVTVYVEEQAAYVEPLENEGIEITETVAPREDNPVEDSQVIVEEVI 300
 QY 301 FVVEEQOEVPETNRTDPEQAKVKKKKPKLLINKFDKTIKALDAEKLRRGKIEBA 360
 DB 301 FVVEEQOEVPETNRTDPEQAKVKKKKPKLLINKFDKTIKALDAEKLRRGKIEBA 360
 QY 361 VNAFKELVRKYPOSPPARYGAQCEEDDLAEKRSRNEVLGALETYQEVASLPDVPADLIX 420
 DB 361 VNAFKELVRKYPOSPPARYGAQCEEDDLAEKRSRNEVLGALETYQEVASLPDVPADLIX 420
 QY 421 LSLIKRSRSDQOFLGHRGSLITLQRLVOLFPPNDTSIKNDLGVGYLLIGNDNAKKYEEV 480
 DB 421 LSLIKRSRSDQOFLGHRGSLITLQRLVOLFPPNDTSIKNDLGVGYLLIGNDNAKKYEEV 480
 QY 481 LSTPNDGFAKYVYFTLKAONKIAESIPLYKEGIESGDPGTDGSRFYFHLGDAMQRYGN 540
 DB 481 LSTPNDGFAKYVYFTLKAONKIAESIPLYKEGIESGDPGTDGSRFYFHLGDAMQRYGN 540
 QY 541 KEAYKMYELGHRGHFASVWQSLVNVNGIKAPMWTPKYGTGTELVKSLERNKILIRDE 600
 DB 541 KEAYKMYELGHRGHFASVWQSLVNVNGIKAPMWTPKYGTGTELVKSLERNKILIRDE 600
 QY 601 GLAVMDKAGFLPREDEENREKGDMSOFTLMQOGRNENNAKCAPYCTLLLEKFPETTCG 660
 DB 601 GLAVMDKAGFLPREDEENREKGDMSOFTLMQOGRNENNAKCAPYCTLLLEKFPETTCG 660

QY 661 REGQIKYSIMBGTWHPHTGPTNCRLNHGLVLPKEGCKIRCANETRTWEGKYLID 720
 DB 661 RRGQIKYSIMBGTWHPHTGPTNCRLNHGLVLPKEGCKIRCANETRTWEGKYLID 720

QY 721 DSFEHEWODASSFRLIFFIVDWMHPELTPOQRSLPAI 758
 DB 721 DSFEHEWODASSFRLIFFIVDWMHPELTPOQRSLPAI 758

RESULT 8

US-09-736-457-1806
 ; Sequence 1806, Application US/09736457
 ; Patent No. US20020168637A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Wang, Tongcong
 ; APPLICANT: Bangur, Chaitanya S.
 ; APPLICANT: Lodes, Michael A.
 ; APPLICANT: Fanger, Gary
 ; APPLICANT: Vedvick, Tom
 ; APPLICANT: Carter, Darrick
 ; APPLICANT: Retter, Marc
 ; APPLICANT: Mannion, Jane
 ; APPLICANT: Fan, Liqun
 ; APPLICANT: Wang, Aijun
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
 ; FILE REFERENCE: 210121.478C15
 ; CURRENT APPLICATION NUMBER: US/09/736,457
 ; NUMBER OF SEQ ID NOS: 2000-12-13
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 1806
 ; LENGTH: 255
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-09-736-457-1806

Query Match 32.8%; Score 1320; DB 9; Length 255;
 Best Local Similarity 98.8%; Pred. No. 1.9e-87;
 Matches 252; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 59 MVALIGWTSVAVVMFPLVJVEEYLGKLGIVDADGDFVDADAKVLLGLKERSTSEPA 118
 DB 1 MVALIGWTSVAVVMFPLVJVEEYLGKLGIVDADGDFVDADAKVLLGLKERSTSEPA 60
 QY 119 VPPEBAEPHTEPEEVPVPAEAPONTEDBAKQIOSLHEMVAHEVEGEDLOQEDGPTEBP 178
 DB 61 VPPEBAEPHTEPEEVPVPAEAPONTEDBAKQIOSLHEMVAHEVEGEDLOQEDGPTEBP 120
 QY 179 PQOEDDEFIATDVDDRFETLEPEVSHETESHVHEETVSQDCNQDMEMNSEQENDPSS 238
 DB 121 PQOEDDEFIATDVDDRFETLEPEVSHETESHVHEETVSQDCNQDMEMNSEQENDPSS 180
 QY 239 SEPVEDERLHHDNDVTVYVEEQAAYVEPLENEGIEITETVAPREDNPVEDSQVIVEEVI 298
 DB 181 SEPVEDERLHHDNDVTVYVEEQAAYVEPLENEGIEITETVAPREDNPVEDSQVIVEEVI 240
 QY 299 SIFPVEEQOEVPET 313
 DB 241 SIFPVEEQOEVPET 255

RESULT 9

US-09-902-941-1806
 ; Sequence 1806, Application US/09902941
 ; Patent No. US20020172952A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Henderson, Robert A.
 ; APPLICANT: Wang, Tongcong
 ; APPLICANT: Watanabe, Yoshihiro
 ; APPLICANT: Johnson, Jeffrey C.
 ; APPLICANT: Retter, Marc W.
 ; APPLICANT: Marnerakis, Margarita

```

; APPLICANT: Carter, Derrick
; APPLICANT: Fanger, Gary R.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: McNabb, Andria
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.478C17
; CURRENT APPLICATION NUMBER: US/09/902,941
; CURRENT FILING DATE: 2001-07-10
; NUMBER OF SEQ ID NOS: 2002
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1806
; LENGTH: 255
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-902-941-1806
```

```
Query Match          32.8%; Score 1320; DB 9; Length 255;
Best Local Similarity 98.8%; Pred. No. 1.9e-87;
Matches 252; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
```

```
QY 59 MVALLGWTSVAVWVMDLVDEYEVGLKGIYDADGDPVDVDAKVLGKERSTSEPA 118
DB 1 MVALLGWTSVAVWVMDLVDEYEVGLKGIYDADGDPVDVDAKVLGKERSTSEPA 60
QY 119 VPPEAPPHTEPPEQVPEAPNPONIEDBAKEQIOSLLHEWVAHVEGEDIQOEDGPTGE 178
DB 61 VPPEAPPHTEPPEQVPEAPNPONIEDBAKEQIOSLLHEWVAHVEGEDIQOEDGPTGE 120
QY 179 POEDDEFMATVDVDRFETLEPEVSHBETESHYVEETVSQDCNQMEEMMSQENPDS 238
DB 121 POEDDEFMATVDVDRFETLEPEVSHBETESHYVEETVSQDCNQMEEMMSQENPDS 180
QY 239 SEPVEDERLHNDTDDVTYQVVEQAVYEPLENGIEITETVAPEDNPVEDSQVIVEEV 298
DB 181 SEPVEDERLHNDTDDVTYQVVEQAVYEPLENGIEITETVAPEDNPVEDSQVIVEEV 240
QY 299 SIFPVEEQOEVPPEPT 313
DB 241 SIFPVEEQOEVPPEPT 255
```

```
RESULT 10
US-09-849-626-1806
; Sequence 1806, Application US/09849626
; Publication No. US20020197669A1
; GENERAL INFORMATION:
; APPLICANT: Bangur, Chaitanya
; APPLICANT: Fanger, Gary
; APPLICANT: Wang, Aijun
; APPLICANT: Wang, Tonglong
; APPLICANT: Switzer, Anne
; APPLICANT: McNeill, Patricia
; APPLICANT: Clapper, Jonathan
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.478C16
; CURRENT APPLICATION NUMBER: US/09/849,626
; CURRENT FILING DATE: 2001-05-03
; NUMBER OF SEQ ID NOS: 1926
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1806
; LENGTH: 255
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-849-626-1806
```

```
Query Match          32.8%; Score 1320; DB 9; Length 255;
Best Local Similarity 98.8%; Pred. No. 1.9e-87;
Matches 252; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
```

```
QY 59 MVALLGWTSVAVWVMDLVDEYEVGLKGIYDADGDPVDVDAKVLGKERSTSEPA 118
```

```
DB 1 MVALLGWTSVAVWVMDLVDEYEVGLKGIYDADGDPVDVDAKVLGKERSTSEPA 60
QY 119 VPPEAPPHTEPPEQVPEAPNPONIEDBAKEQIOSLLHEWVAHVEGEDIQOEDGPTGE 178
DB 61 VPPEAPPHTEPPEQVPEAPNPONIEDBAKEQIOSLLHEWVAHVEGEDIQOEDGPTGE 120
QY 179 POEDDEFMATVDVDRFETLEPEVSHBETESHYVEETVSQDCNQMEEMMSQENPDS 238
DB 121 POEDDEFMATVDVDRFETLEPEVSHBETESHYVEETVSQDCNQMEEMMSQENPDS 180
QY 239 SEPVEDERLHNDTDDVTYQVVEQAVYEPLENGIEITETVAPEDNPVEDSQVIVEEV 298
DB 181 SEPVEDERLHNDTDDVTYQVVEQAVYEPLENGIEITETVAPEDNPVEDSQVIVEEV 240
QY 299 SIFPVEEQOEVPPEPT 313
DB 241 SIFPVEEQOEVPPEPT 255
```

```
RESULT 11
US-10-283-017-1806
; Sequence 1806, Application US/10283017
; Publication No. US20030211510A1
; GENERAL INFORMATION:
; APPLICANT: Henderson, Robert A.
; APPLICANT: Wang, Tonglong
; APPLICANT: Matanabe, Yoshihiro
; APPLICANT: Kalos, Michael D.
; APPLICANT: Sleath, Paul R.
; APPLICANT: Johnson, Jeffrey C.
; APPLICANT: Retter, Marc W.
; APPLICANT: Durham, Margarita
; APPLICANT: Carter, Derrick
; APPLICANT: Fanger, Gary R.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: McNabb, Andria
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.478C20
; CURRENT APPLICATION NUMBER: US/10/283,017
; CURRENT FILING DATE: 2002-10-28
; NUMBER OF SEQ ID NOS: 2157
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1806
; LENGTH: 255
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-283-017-1806
```

```
Query Match          32.8%; Score 1320; DB 12; Length 255;
Best Local Similarity 98.8%; Pred. No. 1.9e-87;
Matches 252; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
```

```
QY 59 MVALLGWTSVAVWVMDLVDEYEVGLKGIYDADGDPVDVDAKVLGKERSTSEPA 118
DB 1 MVALLGWTSVAVWVMDLVDEYEVGLKGIYDADGDPVDVDAKVLGKERSTSEPA 60
QY 119 VPPEAPPHTEPPEQVPEAPNPONIEDBAKEQIOSLLHEWVAHVEGEDIQOEDGPTGE 178
DB 61 VPPEAPPHTEPPEQVPEAPNPONIEDBAKEQIOSLLHEWVAHVEGEDIQOEDGPTGE 120
QY 179 POEDDEFMATVDVDRFETLEPEVSHBETESHYVEETVSQDCNQMEEMMSQENPDS 238
DB 121 POEDDEFMATVDVDRFETLEPEVSHBETESHYVEETVSQDCNQMEEMMSQENPDS 180
QY 239 SEPVEDERLHNDTDDVTYQVVEQAVYEPLENGIEITETVAPEDNPVEDSQVIVEEV 298
DB 181 SEPVEDERLHNDTDDVTYQVVEQAVYEPLENGIEITETVAPEDNPVEDSQVIVEEV 240
QY 299 SIFPVEEQOEVPPEPT 313
DB 241 SIFPVEEQOEVPPEPT 255
```

```
Db      241 SIFPVEEQEVPPDT 255

RESULT 12
US-10-017-754-1806
; Sequence 1806, Application US/10017754
; Publication No. US20030054363A1
; GENERAL INFORMATION:
; APPLICANT: Henderson, Robert A.
; APPLICANT: Wang, Tongtong
; APPLICANT: Matanabe, Yoshihiro
; APPLICANT: Johnson, Jeffrey C.
; APPLICANT: Retter, Marc W.
; APPLICANT: Marnetakis, Margarita
; APPLICANT: Carter, Darick
; APPLICANT: Fanger, Gary R.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: McNabb, Andria
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.478C18
; CURRENT APPLICATION NUMBER: US/10/017,754
; CURRENT FILING DATE: 2001-10-29
; NUMBER OF SEQ ID NOS: 2004
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1806
; LENGTH: 255
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-017-754-1806

Query Match      32.8%; Score 1320; DB 14; Length 255;
Best Local Similarity 98.8%; Pred. No. 1.9e-87;
Matches 252; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY      59 MVALIGVWTSVAVWFMDLVYEEVLGKIGIYDADGDGDFVDVDAKVLGLKERSISEPA 118
; GENERAL INFORMATION:
; APPLICANT: OKURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 8153
; LENGTH: 250
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-8153

Db      61 VPPEEAPHTPEEQVVEAPONTIDEAKEQIQLLHEVVAHEHVEGDLQOEDGPTGE 120
QY      119 VPPEEAPHTPEEQVVEAPONTIDEAKEQIQLLHEVVAHEHVEGDLQOEDGPTGE 178
; GENERAL INFORMATION:
; APPLICANT: OKURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 8153
; LENGTH: 250
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-8153

Db      179 POEDEDEFLMATVDVDRFETLEPEVSHETESHVHEETVSQDCNODMEMMSEQENPDS 238
QY      121 POEDEDEFLMATVDVDRFETLEPEVSHETESHVHEETVSQDCNODMEMMSEQENPDS 180
; GENERAL INFORMATION:
; APPLICANT: OKURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 8153
; LENGTH: 250
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-8153

Db      239 SEPVEDERLHHDTDVTVQVVEQAVPEPLENEGIEITETVTAPEPDNPVEDSQVIVEEV 298
QY      181 SEPVEDERLHHDTDVTVQVVEQAVPEPLENEGIEITETVTAPEPDNPVEDSQVIVEEV 240
; GENERAL INFORMATION:
; APPLICANT: OKURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 8153
; LENGTH: 250
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-8153

Db      299 SIFPVEEQEVPPDT 313
QY      241 SIFPVEEQEVPPDT 255

RESULT 13
US-10-113-872-1806
; Sequence 1806, Application US/1011872
; Publication No. US20030170255A1
; GENERAL INFORMATION:
; APPLICANT: Matanabe, Yoshihiro
; APPLICANT: Henderson, Robert A.
; APPLICANT: Kato, Michael D.
; APPLICANT: Sleath, Paul R.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Carter, Darick
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.478C19
; CURRENT APPLICATION NUMBER: US/10/113,872
```

```
; CURRENT FILING DATE: 2002-03-28
; NUMBER OF SEQ ID NOS: 2011
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1806
; LENGTH: 255
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-113-872-1806

Query Match      32.8%; Score 1320; DB 14; Length 255;
Best Local Similarity 98.8%; Pred. No. 1.9e-87;
Matches 252; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY      59 MVALIGVWTSVAVWFMDLVYEEVLGKIGIYDADGDGDFVDVDAKVLGLKERSISEPA 118
; GENERAL INFORMATION:
; APPLICANT: OKURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 8153
; LENGTH: 250
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-8153

Db      119 VPPEEAPHTPEEQVVEAPONTIDEAKEQIQLLHEVVAHEHVEGDLQOEDGPTGE 178
QY      61 VPPEEAPHTPEEQVVEAPONTIDEAKEQIQLLHEVVAHEHVEGDLQOEDGPTGE 120
; GENERAL INFORMATION:
; APPLICANT: OKURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 8153
; LENGTH: 250
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-8153

Db      179 POEDEDEFLMATVDVDRFETLEPEVSHETESHVHEETVSQDCNODMEMMSEQENPDS 238
QY      121 POEDEDEFLMATVDVDRFETLEPEVSHETESHVHEETVSQDCNODMEMMSEQENPDS 180
; GENERAL INFORMATION:
; APPLICANT: OKURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 8153
; LENGTH: 250
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-8153

Db      239 SEPVEDERLHHDTDVTVQVVEQAVPEPLENEGIEITETVTAPEPDNPVEDSQVIVEEV 298
QY      181 SEPVEDERLHHDTDVTVQVVEQAVPEPLENEGIEITETVTAPEPDNPVEDSQVIVEEV 240
; GENERAL INFORMATION:
; APPLICANT: OKURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 8153
; LENGTH: 250
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-8153

Db      299 SIFPVEEQEVPPDT 313
QY      241 SIFPVEEQEVPPDT 255

RESULT 14
US-10-156-761-8153
; Sequence 8153, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OKURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 8153
; LENGTH: 250
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-8153

Query Match      6.8%; Score 275; DB 14; Length 250;
Best Local Similarity 33.3%; Pred. No. 9.3e-12;
Matches 65; Conservative 31; Mismatches 91; Indels 8; Gaps 4;
```

Qy 684 NCRIMHIGLVIPKGGKIRCANETRWEEGKVLIFDDSEHEVWQDASSRLIFIVDW 743
Db 166 NFSINLHLAVDIP-DGCSITVAGETRSWEEGKCLLFYSPFEHARNTGTPTCLLIDW 224
Qy 744 HPELTPOQRSLPAI 758
Db 225 HPELTVPERRALVAL 239

RESULT 15

US-10-168-274-15
; Sequence 15, Application US/10168274
; Publication No. US20030124106A1
; GENERAL INFORMATION:
; APPLICANT: INCYTE GENOMICS, INC.
; APPLICANT: YUE, Henry
; APPLICANT: LAL, Preeti
; APPLICANT: TANG, Y. Tom
; APPLICANT: HILLMAN, Jennifer
; APPLICANT: BAUGHN, Mariah R.
; APPLICANT: AZIMZAI, Yalda
; APPLICANT: LU, Dzung Aina M.
; TITLE OF INVENTION: HUMAN OXIDOREDUCTASE PROTEINS
; FILE REFERENCE: PP-0754 PCT
; CURRENT APPLICATION NUMBER: US/10/168,274
; CURRENT FILING DATE: 2002-08-26
; PRIOR APPLICATION NUMBER: 60/172,367
; PRIOR FILING DATE: 1999-12-16
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PERL Program
; SEQ ID NO 15
; LENGTH: 369
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20030124106A1 2754425CD1
US-10-168-274-15

Query Match

6.8%; Score 274.5; DB 14; Length 369;

Best Local Similarity 30.1%; Pred. No. 1,7e-11; Mismatches 105; Indels 25; Gaps 9;

Matches 75; Conservative 44; Mismatches 105; Indels 25; Gaps 9;

Qy 526 RFFYHLDAMQRYGNKEAYKMYELG-HK---RGHF---ASYWQRSLYNVNGLKAQPW 575
Db 125 KLYHNLDQYAKR-----YWSGWMGRIGHGIRBQGRYINSRPSIQKPEVFFLPDLPPTPY 178

Qy 576 WTPKETGYTELVSLEBNWKLIRDEGLAVNDKAKGLFLPR--DENLAEKGDWQFTLMQ 633
Db 179 FSRDAQHD--VEVLEHNFOTLICEFETLYKAFSNCSLPQGWKXNSTPSSGEMFTFYLVNQ 236

Qy 634 GRNENACKGAPKTCILLEKFPETTGR-RGQIKYSIMHPTGVHWPHTGPTNCRRLMHLG 692
Db 237 GVCVPRNCRKCPRTYRLIGSLRTIGNNVFNACISVLSPTVITEHYGPTNIRIRCHLG 236

Qy 693 LVIPKGGCKIRCANETRWEEGKVLIFDDSEHEVWQDASSF---RLIFIVDWHPELTP 749
Db 297 LKTP-NCCELVVGGEPCWMAEGRCILFDDSFLLHAFFHESGAEDGPRVVFVVDLMHPVAA 355

Qy 750 QQRSLPAI 758
Db 356 AERQALDPI 364

Search completed: May 19, 2004, 16:09:34
Job time : 76.9354 secs

Gencore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 19, 2004, 15:54:54 ; Search time 31.7034 Seconds

(without alignments)
2299.852 Million cell updates/sec

Title: US-09-903-199-2

Perfect score: 4022
Sequence: 1 MAQRKNAKSSGNSSSSSGSSGSS.....YDVWHPRLTPQGRSLPAL 758Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 28366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 28366

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: PIR1:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3966.5	98.6	757	138423	aspartyl beta-hydr
2	3172	78.9	754	1 BABOH	peptide-aspartate
3	890	22.1	270	3 C7792	cardiac junctate-1
4	798	19.8	872	2 T1861	probable peptide-a
5	263.5	6.6	186	2 T47148	hypothetical prote
6	216.5	5.4	1110	2 T51116	NF-180 - sea lamp
7	199	4.9	1271	2 A4555	glutamate rich pro
8	191.5	4.8	312	2 H83527	hypothetical prote
9	197.5	4.8	1881	2 H95076	zinc metalloprotei
10	189.5	4.7	1616	2 G64242	cytochrome-acces
11	188.5	4.7	722	2 T42963	hypothetical prote
12	187	4.6	706	2 A45990	junctional sarcopl
13	183	4.5	763	2 T08929	hypothetical prote
14	181	4.5	771	1 A33430	h-caldesmon - chic
15	179.5	4.5	411	2 S47436	flagellar antigen
16	178.5	4.4	729	2 S68191	titadin - human
17	177	4.4	301	2 A82601	aspartyl/asparagin
18	176	4.4	845	2 A45669	neurofilament trip
19	177	4.4	1094	2 S49313	protein kinase - s
20	174.5	4.3	1948	2 S00485	gene 11-1 protein
21	173.5	4.3	1871	2 D96796	probable heat shoc
22	172.5	4.3	1240	2 S52734	hypothetical prote
23	172	4.3	465	2 S46759	hypothetical prote
24	172	4.3	630	2 S29786	hypothetical prote
25	171.5	4.3	607	2 S27776	60K protein (allel
26	171	4.3	1804	2 T34518	nestin - golden ha
27	169.5	4.2	1233	2 S56271	lamin B2 - mouse
28	169.5	4.2	1233	2 S56271	hypothetical prote
29	168.5	4.2	695	2 T40168	hypothetical prote

30	168.5	4.2	913	2 T52485	neurofilament prot
31	168	4.2	302	2 AF1021	probable membrane-
32	168	4.2	1192	2 A71623	probable secreted
33	167.5	4.2	1876	2 E97942	zinc metalloprotei
34	166.5	4.1	1624	2 T25592	hypothetical prote
35	165	4.1	651	2 S18874	hypothetical prote
36	164.5	4.1	7962	2 T38346	nucleolin - Africa
37	164	4.1	1957	2 T38077	elastic titin - hu
38	163.5	4.1	1957	2 T38077	hypothetical coile
39	163	4.1	1132	2 T43483	translational initia
40	163	4.1	501	2 C71948	peptidylprolyl iso
41	163	4.1	1877	2 T21861	hypothetical prote
42	163	4.1	1939	1 S21801	hypothetical prote
43	162.5	4.0	299	1 B83081	myosin heavy chain
44	162.5	4.0	793	1 JH0628	hypothetical prote
45	162.5	4.0	992	2 T46337	caldesmon - human
					hypothetical prote

ALIGNMENTS

RESULT 1	
138423	
aspartyl beta-hydroxylase - human	
C:Species: Homo sapiens (man)	
C:Date: 29-May-1998 #sequence_revision 29-May-1998 #text_change 21-Jan-2000	
C:Accession: I38423	
R:Korlosh, F.; Gieffers, C.; Frey, J.	
Gene 150, 395-399, 1994	
A:Title: Cloning and characterization of the human gene encoding aspartyl beta-hydroxyl	
A:Reference number: I38423; PMID:95121937; PMID:7821814	
A:Accession: I38423	
A>Status: preliminary; translated from GB/EMBL/DDBJ	
A:Molecule type: mRNA	
A:Residues: 1-757 <RES>	
A:Cross-references: EMBL:U03109; NID:G458031; PID:AA82108.1; PID:G458032	
C:Superfamily: peptide-aspartate beta-dioxygenase; tetrairicopeptide repeat homology	
F:54-75/Domain: transmembrane #status predicted <TRM>	
F:341-374/Domain: tetrairicopeptide repeat homology <TR>	
Query Match	
Best local Similarity 99.2%; Pred. No. 1e-195; 3; Indels 1; Gaps 1;	
Matches 752; Conservative 2; Mismatches 3; Indels 1; Gaps 1;	
QY	1 MAQRKNAKSSGNSSSGSGSGSTAGSSSPGARRETKGQKXKRGKGLSGTSFTWFMV 60
DB	1 MAQRKNAKSSGNSSSGSGSGSTAGSSSPGARRETKGQKXKRGKGLSGTSFTWFMV 60
QY	61 IALLGWTSVAVWVFDLVNVEVTLGKLGIVDADGDDFDVDDAKVLGKERSTSEAVP 120
DB	61 IALLGWTSVAVWVFDLVNVEVTLGKLGIVDADGDDFDVDDAKVLGKERSTSEAVP 120
QY	121 PEABPHTPEPEQVPVBAEPQNTIEDBAKEIQSLHEMVAAHEVEGEDLQOEDGPTGPEPQ 180
DB	121 PEABPHTPEPEQVPVBAEPQNTIEDBAKEIQSLHEMVAAHEVEGEDLQOEDGPTGPEPQ 180
QY	122 PEABPHTPEPEQVPVBAEPQNTIEDBAKEIQSLHEMVAAHEVEGEDLQOEDGPTGPEPQ 180
DB	122 PEABPHTPEPEQVPVBAEPQNTIEDBAKEIQSLHEMVAAHEVEGEDLQOEDGPTGPEPQ 180
QY	181 QEDDEFLMATVDVDRFETLEPEVSHETESHVVEETVSQDCQDMEMMSSEQNDPSSE 240
DB	181 QEDDEFLMATVDVDRFETLEPEVSHETESHVVEETVSQDCQDMEMMSSEQNDPSSE 240
QY	241 PVVEDERLHHDDVDVTVQVVEBAVVEPLENBEIGETEVAPAPEDNPVEDSOVVEEVS 300
DB	241 PVVEDERLHHDDVDVTVQVVEBAVVEPLENBEIGETEVAPAPEDNPVEDSOVVEEVS 300
QY	301 FPVEEQEVPETNRKTIDPEQAKYKKKKPKLNFKEKTIKALDAEKLRRKGIIEBA 360
DB	301 FPVEEQEVPETNRKTIDPEQAKYKKKKPKLNFKEKTIKALDAEKLRRKGIIEBA 360
QY	361 VNAFKELVRRYPOSPPARYKAAQCEDDLAEKRSNEVLRGAIFTYQVVASIPVPAULK 420
DB	361 VNAFKELVRRYPOSPPARYKAAQCEDDLAEKRSNEVLRGAIFTYQVVASIPVPAULK 420
QY	421 LSLKRRSDRQQLGHRGSLTLTQRLVQLFPNDTSLKNDGVGYLLGDNDAKAYEEV 480


```

Db 421 LSKRRSDRQOFLGMRSLTLTLOQLVQLFPNDTSLKNDLGVGLIGDNNNAKVEEV 480
Qy 491 LSVTPNDGFAYKHVFFLLKAQNKIAESIPLYKEGIESGDPGTDDGRFFYHLGDAMORVN 540
Db 491 LSVTPNDGFAYKHVFFLLKAQNKIAESIPLYKEGIESGDPGTDDGRFFYHLGDAMORVN 540
Qy 541 KEATKWEVLGHRKHGFASVWQSRSLYVNVGLKAQCPMTPTKETGYELVKSLENNKILRDE 600
Db 541 KEATKWEVLGHRKHGFASVWQSRSLYVNVGLKAQCPMTPTKETGYELVKSLENNKILRDE 599
Qy 601 GLAVMDKXGLFLPEDEVLREKGDMSQFTLMQOGRNNAKGA PKCTCTTLEKEPPTTGC 660
Db 600 GLAVMDKXGLFLPEDEVLREKGDMSQFTLMQOGRNNAKGA PKCTCTTLEKEPPTTGC 659
Qy 661 RRGQIKYSIMHPTGTHVPTGTCRLRMHLGLVLPKEGCKIRCANETRTWEEGKVLIFD 720
Db 660 RRGQIKYSIMHPTGTHVPTGTCRLRMHLGLVLPKEGCKIRCANETRTWEEGKVLIFD 719
Qy 721 DSFEHEWQDASSFRLLIFVDVWHELTTPQQRSLPAI 758
Db 720 DSFEHEWQDASSFRLLIFVDVWHELTTPQQRSLPAI 757

```

RESULT 2

```

BABOH
peptide-aspartate beta-dioxygenase (EC 1.14.11.16) - bovine
N.Alternate names: aspartyl (asparaginyl) beta-hydroxylase
C.Species: Bos primigenius taurus (cattle)
C.Date: 31-Dec-1993 #sequence revision 10-40-1995 #text_change 11-Jun-1999
A.Accession: A42969; A39470; B39470; C39470; S27948
R.Title: S. Vanuise, W.J.; Dien, R.E.; Kohl, N.E.; Dixon, R.A.; Elliston, K.O.; Stern,
J. Biol. Chem. 267, 14322-14327, 1992
A.Title: cDNA cloning and expression of bovine aspartyl (asparaginyl) beta-hydroxylase.
A.Reference number: A42969; MUID:92332546; PMID:1378441
A.Accession: A42969
A.Molecule type: mRNA
A.Residues: 1-754 <JIA>
A.Cross-references: EMBL:M91213; NID:G162693; PIDN:AAA03563.1; PID:G162694
A.Experimental source: brain
A.Note: sequence extracted from NCBI backbone (NCBI:108534)
R.Mang, Q.; Vandusen, W.J.; Petroski, C.J.; Gareky, V.M.; Stern, A.M.; Friedman, P.A.
J. Biol. Chem. 266, 14004-14010, 1991
A.Title: Bovine liver aspartyl beta-hydroxylase. Purification and characterization.
A.Reference number: A39470; MUID:91310689; PMID:1856229
A.Accession: A39470
A.Molecule type: protein
A.Residues: 289-328 <MAN>
A.Accession: B39470
A.Molecule type: protein
A.Residues: 615, 'X', 617-630, 'XX', 633-634, 'X', 636, 'XX', 639-641 <MA2>
A.Accession: C39470
A.Molecule type: protein
C.Comment: This enzyme uses ferrous iron as a cofactor, and while beta-hydroxylating the
C.Comment: Aspartic acid and asparagine residues in the Bgf homology domain of certain P
C/Keywords: glycoprotein; oxidoreductase; transmembrane protein
F.57-78/Domain: intracellular #status predicted <INCL>
F.57-78/Domain: transmembrane #status predicted <TRM>
F.289-754/Product: peptide-aspartate beta-dioxygenase, 56K form #status predicted <56K>
F.337-370/Product: peptide-aspartate beta-dioxygenase, 52K form #status predicted <52K>
F.337-370/Domain: tetrahydrocortisol repeat homology <THL>
F.337-404/Domain: tetrahydrocortisol repeat homology <THL>
F.13, 96, 466, 702/Binding site: carbohydrate (Asn) (covalent) #status predicted

```

Query Match 78.9% Score 3172; DB 1; Length 754;

Best Local Similarity 78.7% Pred. No. 4, 8e-15; Indels 44; Gaps 6;

Matches 612; Conservative 45; Mismatches 77;

```

Qy 1 MAQRNKAASG---NSSSSGSGSGSTASGSSPGARRPTKHGKNGKXGSLSGTSFTTW 57
Db 1 MAPRNKAAGGNGSSSSSGSGSPTGCTSGGSSSPGARRETKOGLKNGKXGSLSGTSFTTW 60

```

```

Qy 58 FMVIALGVMTSAVAVFVFDLVDEEVL-----GKLTIDADGDGDFVDD 102
Db 61 FMVIALGVMTSAVAVFVFDLVDEEVLAKAKDFRNVNLSVILQGLGTYDADGDGDFVDD 120
Qy 103 AKVLLGKKEKSTGEPAVPEEAPHTPEEQQVVEAPQVIEDBAKQIQSLHEVNAE 162
Db 121 AKVLLGKKEKSTGEPAVPEEAPHTPEEQQVVEAPQVIEDBAKQIQSLHEVNAE 179
Qy 163 HVEGEDVQDDEGPTGEQDDEFLMATVDVDDRETLEPEVSHETSYSVEETVSQDC 222
Db 180 --PGENIPQPEPGEAELODDHVFVSDADDREKPGTAGVHEETDSHIEETASAY 237
Qy 223 NQDEEMMSQENDDSEPEVED--ERLHNDTDVTVQVVEQAVPELENEGIEITVT 280
Db 238 SQDMEWVQENDDSEPEVVDABETVQETDVTYRDYDEQ----- 280
Qy 281 APPEDNVEVSQVYVEVVSFPVEEQDVEPPTNKRTDDPEQAKVKKKKKLINKPDKT 340
Db 281 ---DHAVDNNTLIEPHEPPEAEQVEPPTNKKADEPKKGVKKKKKLINKPDKT 336
Qy 341 IKAELDAEKLRRKGTSEAVNAFKELVRYKYPQSPARVYKACEDDLAEKRSNEVLRG 400
Db 337 IKAELDAEKLRRKGTSEAVNAFKELVRYKYPQSPARVYKACEDDLAEKRSNEVLRR 396
Qy 401 ALETYQVAVSLPVVADLKLILKRRSDRQOFLGMRSLTLTLOQLVQLFPNDTSLKNDL 460
Db 397 ALETYQVAVSLPVVADLKLILKRRSDRQOFLGMRSLTLTLOQLVQLFPNDTSLKNDL 456
Qy 461 GVGGLLIGDNDNAKKVVEVLTVPNDGFKVHYGFLTXKONKIAESIPLYKEGIESGDP 520
Db 457 GVGGLLIGDNDNAKKVVEVLTVPNDGFKVHYGFLTXKONKIAESIPLYKEGIESGDP 516
Qy 521 GTDDGRFFYHLGDAMQVRVGNKEAVKWEVLGHRKHGFASVWQSRSLYVNVGLKAQCPMTPEKE 580
Db 517 GTDDGRFFYHLGDAMQVRVGNKEAVKWEVLGHRKHGFASVWQSRSLYVNVGLKAQCPMTPEKE 576
Qy 581 TGYTELKYSLEIRWKLIRDEGLAVMDKXGLFLPEDEVLREKGDMSQFTLMQOGRNNA 640
Db 577 TGYTELKYSLEIRWKLIRDEGLAVMDKXGLFLPEDEVLREKGDMSQFTLMQOGRNNA 636
Qy 641 CKGAPKCTTLEKEPPTTGGRRGQIKYSIMHPTGTHVPTGTCRLRMHLGLVLPKEGC 700
Db 637 CKGAPKCTTLEKEPPTTGGRRGQIKYSIMHPTGTHVPTGTCRLRMHLGLVLPKEGC 696
Qy 701 KIRCANETRTWEEGKVLIFDSEHEWQDASSFRLLIFVDVWHELTTPQQRSLPAI 758
Db 697 KIRCANETRTWEEGKVLIFDSEHEWQDASSFRLLIFVDVWHELTTPQQRSLPAI 754

```

RESULT 3

```

JC7792
cardiac junctate-1 - mouse
C.Species: Mus musculus (house mouse)
C.Date: 02-Apr-2002 #sequence_revision 02-Apr-2002 #text_change 03-May-2002
A.Accession: JC7792
R.Hong, C.S.; Kwak, Y.G.; Ji, J.H.; Chae, S.W.; Kim, D.H.
Biochem. Biophys. Res. Commun. 289, 882-887, 2001
A.Title: Molecular cloning and characterization of mouse cardiac junctate isoforms.
A.Reference number: JC7792
A.Accession: JC7792
A.Contents: Heart
A.Molecule type: mRNA
A.Residues: 1-270 <HON>
A.Cross-references: GB:AF302653
C.Comment: This protein, a Ca2+ binding protein, plays a role both in contractile and c
C/Keywords: cardiac muscle, heart

```

Query Match 22.1% Score 890; DB 2; Length 270;

Best Local Similarity 63.9% Pred. No. 8, 8e-33; Indels 14; Gaps 5;

Matches 179; Conservative 30; Mismatches 57;

```

Qy 34 RETGKHGKNGKRGKGLSGTSFTTWFMVIALGVMTSAVAVFVFDLVDEEVLGKLTIDAD 93

```

Db 5 KEAKHGKNGKRGSGSGSFFTWFWVIALLGWTVAAVWVFDLVYEVEVLGKLGYYDAD 64
QY 94 GDDDFDVDDAKVLLGLKERSTSEPAVPEAEHPTEPEQVPEABPQNIIEDEAKXQIQIS 153
Db 65 GDDDFDVDDAKVLLGLKERSPERTPP-BAETHALEQAPFAGAIQVNEBEVKQIQIS 123
QY 154 LHEWVAHEVEGEDLQOEDGPTGPEQOEDFEIATDVDDRFETLEPEVSHETESHYH 213
Db 124 LLOESVHTDH---DL-BADGLAGEPQPEVEDFLTYTDDSDRFEDLEPGTVHEIEDTYH 178
QY 214 VETTVQDCNQDMEEMSEQENDPSEPVYEDERLHHTDDVTYYQVEQAYEPIENEG 273
Db 179 VEDTASQNHNDMEETNEQENDPSEAVTDAGVLLPHAEVYHQDYDE-PVYEPSEHEG 237
QY 274 IETETVAPPEPDVPEDSQVIVEEVSIFPYEEQOEVPEPT 313
Db 238 VEIS-----DNTIDDSITISEIIVASVEEQDTPPDT 270

RESULT 4

T18861
probable peptidase-aspartate beta-dioxygenase (BC 1.14.11.16) - *Caenorhabditis elegans*
C:Species: *Caenorhabditis elegans*
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
C:Accession: T18861, T23533
R:Swindburne, J.
submitted to the EMBL Data Library, August 1996
A:Reference number: Z19032
A:Accession: T18861
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-872 <WIL>
A:Cross-references: EMBL:Z79596; NID:e1323798; PIDD:CA01859.1; GSPDB:GN00028; CESP:K09A
A:Experimental source: clone C0206
R:Swindburne, J.
submitted to the EMBL Data Library, August 1996
A:Reference number: Z19753
A:Accession: T23533
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-872 <WIL>
A:Cross-references: EMBL:Z79601; PIDD:CA01897.1; GSPDB:GN00028; CESP:K09A9.6
A:Experimental source: clone K09A9
C:Genetics:
A:Gene: CESP:K09A9.6
A:Map position: X
A:Introns: 18/1; 40/3; 74/1; 104/2; 139/1; 170/3; 385/3; 432/2; 464/3; 530/2; 609/3; 776
C:Keywords: oxidoreductase

Query Match 19.8%; Score 798; DB 2; Length 872;
Best Local Similarity 26.8%; Pred. No. 2e-33;
Matches 191; Conservative 144; Mismatches 278; Indels 100; Gaps 15;
QY 89 TTDADGDFVDVDAKVLGLKERSTSEPAVPEAEHPTEPEQ-----VPYAEFQ 141
Db 205 VEDDDDDDDDDDDDDVVE-----APAEQERSKQKANKHREKKKKYQPYAEED 255
QY 142 NIEDEAKEQIQSLHEWVAHEVEGEDLQOEDG-----PTGPEQOEDFEIATDVDDR 195
Db 256 LDDDDDDDDDDDDGDDDDDEKDGNDAKDKDGGDDDDDDDDDDDDDDDEGKKKKTSVEAK 315
QY 196 FETL-----EPEVSHETE--HSYVEET-----VSQDC-NQDMEE- 228
Db 316 SKVVEEKNDGDEPHVSHREAOQLRKRLRSRDNRRQAGNRECIHQDCPNRESLKP 375
QY 229 ---MMSEQNDPDSSEPVVEDERLHHTDDVTYYQVEQAYEPIENEGIEITETVAPPE 284
Db 376 RKSLLVTKKTKTSAMVERIIDDDEDDDDDD-----EDDSEVKN-----N 415
QY 285 DKPVEDSQVIVEEVSIFPYEEQOEVPEPTNRTDDEQAKYKKKKKPLINKFDKTIKAE 344
Db 416 DREDDDDDDVDERIS-----DRSSSSYKHAITTKKEIGFRDI 454

QY 345 IDAAEKLRKRGKIEEAVNAFKELVYKRPQSPRPRYKQACEDDIAEKRRSNEVLRGAIET 404
Db 455 LDRANLVKEHQYEEAMBLFDHVIAYPASTATYKGAARAYIRGHEIDEDDRDAIETI 514
QY 405 YQEVASLPDVPADLLKLSIKRRSDROQFLGHRGSLITLQRLVOLFPNDTSLKNDLGIVY 464
Db 515 YKILQNSGVDPALPRQAARLIETKRFRGQJHKLTTHARVFLDRPEELINQTDPAISF 574
QY 466 LLIGNDNAKRYEEVLSTPNDGFAKHYGFIKA-QNKIASIPYLMKEGIESGPGND 523
Db 575 VMKKYEADRYTLKVNLANDPNHTVIALAYYGILKAHDKVQGVALMKRSKLNADNEIT 634
QY 524 DDFEFYGLDAMQVRGNK-EAYKWEYELGHRGHFASVWORSLYNVNGLAOPWTPKETE 582
Db 635 DKFETYLQGHGTLTGKRSSEADAVYQKAQMVFMTAQORSLYNIGLGRAMWMDQTP 694
QY 583 YTELVSLENNKILRDGLAVMDKAGLFLPEDENLRKSGPMSOFTLMQOGRNENACK 642
Db 695 YSKFLKTVERQWATIROEGMEVLKDCSDCWLHNOQLVIDGQWKFPIMSEQNFIKSCS 754
QY 643 GAKTCITLLEKPEPTTGCRRGOIKYSIMHPGTHVPHPTGTCRLRMHGLVTPKESGCT 702
Db 755 RWPQTCITLQEPAASSNASKSDMHLSTVSSGASILPHCPPTNYHLQAHGLVSPS-ARI 813
QY 703 RCANETRTMBGKVLIFDDSFHEVWOD---ASSFRLIFIVVWHPETLPOOR 752
Db 814 RVGNLTGKMRSGKFLTYDSSFHEHLOFGDASSSFRVLTITLDMHEVQPHQR 866

RESULT 5

T47148
hypothetical protein DKFZp761P039.1 - human (fragment)
C:Species: *Homo sapiens* (man)
C>Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000
C:Accession: T47148
R:Blum, H.; Bauersachs, S.; Mewes, H.W.; Weil, B.; Wiemann, S.
submitted to the Protein Sequence Database, March 2000
A:Reference number: Z24379
A:Accession: T47148
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-186 <AAA>
A:Cross-references: EMBL:AL161993
A:Experimental source: adult amygdala; clone DKFZp761P039
C:Genetics:
A:Note: DKFZp761P039.1

Query Match 6.6%; Score 263.5; DB 2; Length 186;
Best Local Similarity 35.4%; Pred. No. 6.5e-07;
Matches 63; Conservative 28; Mismatches 80; Indels 7; Gaps 4;
QY 587 VKSLERNKTLRDGLAVMDKAGLFLPE--DENLRKSGDMSOFTLMQOGRNENACKA 644
Db 5 VEVLERNPQTLICETPTIKAFNSCLPQGMKMNSTPSGEWTFVLVNGGVCPNCRKC 64
QY 645 PXTCTLLEKPEPTTGCRRGOIKYSIMHPGTHVPHPTGTCRLRMHGLVTPKESGCT 703
Db 65 PRTYLLDSLRICLNNVFNAGACISVLSPGTYITHTGYTNIRICHGLKTP-NGCELV 123
QY 704 CANETRTMBGKVLIFDDSFHEHWDASSF---RLIFVYVWHPETLPOQRSPLAI 758
Db 124 VQGEQCAWABRGCLIFDDSFHAAFHGSAEDGPRVFWVDLHHPVAAERQALDFI 181

RESULT 6

151116
NF-180 - sea lamprey
C:Species: *Petromyzon marinus* (sea lamprey)
C>Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 21-Jul-2000
C:Accession: I51116
R:Jacobs, A.J.; Kamboj, J.; Seitzer, M.E.
Brain Res. Mol. Brain Res. 29, 43-52, 1995

QY 702 IRCANETWEBSKULIFDDSEFEHYNQDASSPRLIFIVVWHP 745
DB 185 IYVDGQPYAWRDGEDVMEDEFVHWVKNTEQTRVILFCDIERP 228

RESULT 9

H95076

C:Species: Streptococcus pneumoniae
C:Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 03-Aug-2001

C:Accession: H95076

R:Reteller: H.; Nelson, K.E.; Paulsen, J.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heid

son, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapple,

Science 293, 498-506, 2001

A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,

A:Title: Complete Genome Sequence of a Virulent isolate of Streptococcus pneumoniae.

A:Reference number: A95000; MUID:21357209; PMID:11463916

A:Accession: H95076

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1881 <KUR>

A:Cross-references: GB:AE005672; PIDN:AAK74809.1; PID:G14972138; GSPDB:GM00164; TIGR:SP4

A:Experimental source: strain TIGR4

C:Genetics: SP0664

Query Match 4.8%; Score 191.5; DB 2; Length 1881;

Best Local Similarity 21.6%; Pred. No. 0.06;

Matches 138; Conservative 82; Mismatches 226; Indels 193; Gaps 30;

QY 111 ERSTSEPAVPER-APHTPEBOQVPEAEAPONIDEAKQIQLHEMVAHEVGED 168
DB 211 KEDSABPA-FVEEVGGEVSKPEKVAVKESQPSKPAESKYEQAGEPV---APRED 265
QY 169 LQGEDGTGEPQOEDDEFLMATVDVDRFETLEBEVSHETHEHYNHEVTSQDQCNQ-DME 227
DB 266 EKAPVEPEKQPEAPEE--KAVEETPKQESSTPDYAEETVER--KEEIVNQSIQPKHE 321
QY 228 EMMSBOENPDSSPEVDE---RLHHDVDTVQYVEEQAVYEPLENEGIEETEYVAP 283
DB 322 TPAREKQTEPTEEPKVEQAGEVAPAREDEQAPAPAREKQPEVPEEKAV---EETPKP 378
QY 284 ED-----NPVEDSQV---LYEVSIFPVE-----EQQ 307
DB 379 EDKIKGIGTKEPVDKSELNNQIDKASSVPTDYSTASYNALGVLLETAKGVASPEVKQ 438
QY 308 EEPETNR-----KTD-----DPEOKAVYKKKKPKLLNKFDKTIKALDAEK 350
DB 439 EVASSTNKLTALDALNVDTKELNNTIADAKTVKEHYSRDSWQNLQTEVTKAEKVAYANT 498
QY 351 IARRKGLIEAVNAF-----KELVRYPOSPPARVYKQAQ 383
DB 499 DAKQSEVNAEVEKLTATIEKLVLESEKPILTSTDKILEREAVAKTYLE---NQNKIK 555
QY 384 CEDDLAEKRSNVEVLAGAT-----ETVQ-----E 407
DB 556 IKSTITAEKKGEEVINTVVLTDKVTETITSAAFKMLEYKEYTLSTTMYDRGNGEETE 615
QY 408 VASLPVPADLLKLSK--RRSDROOFLGMRGSLTLQRLVQLPFPDSTLKNLDIGVYL 465
DB 616 TLENQNIQLDLKVELKNIKRTDLIKY---ENGKTESLITTIIPDKSNY-----YL 665
QY 466 LIGDNDN-----AKVVEEVLSTPNDGPAKVHYGILK-AONKIAS--IPYIKGIGIES 517
DB 666 KITSNQKQTLTAVKNIIEFTVNGTPYKVTALADNLVSRADNKKPEEYVHYIEK---- 721
QY 518 GDEGTDDGRFYFLGDAMQEVGNKEAVKVELGHKRGHPASVQORSLYNY--NGLKAQPM 575
DB 722 --EKVEHDNYVYFKELVEAIQN--DSKSEYRLGQ-----SMSARNVVPNG---KSY 766
QY 576 WPKETGYTELTVKSLERNMKLIRDEG--LAVMDKAKGLF 612

DB 767 ITKEFTG-----KLISSEKQOPATIELEHPLF 793

RESULT 10

G64242

Cytedherence-accessory protein (hmw1) homolog MG386 - Mycoplasma genitalium

C:Species: Mycoplasma genitalium

C:Date: 17-Nov-1995 #sequence_revision 17-Nov-1995 #text_change 07-Dec-1999

C:Accession: G64242

R:Fraser, C.M.; Gocayne, J.D.; White, O.; Adams, M.D.; Clayton, R.A.; Fleischmann, R.D.

M.; Fuhlmann, J.; Nguyen, D.; Uitterback, T.R.; Saudek, D.M.; Phillips, C.A.; Merrick, J

C.A.; Venter, J.C

Science 270, 397-403, 1995

A:Title: The minimal gene complement of Mycoplasma genitalium.

A:Reference number: A64200; MUID:96026346; PMID:756993

A:Accession: G64242

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-1616 <TIGR>

A:Cross-references: GB:U39723; GB:L43967; NID:G1046092; PID:G1046097; TIGR:MG386

A:Experimental source: strain G-37

C:Genetics:

A:Genetic code: SGC3

Query Match 4.7%; Score 189.5; DB 2; Length 1616;

Best Local Similarity 22.3%; Pred. No. 0.063;

Matches 142; Conservative 78; Mismatches 205; Indels 213; Gaps 35;

QY 75 FDLVDVDEVLGK-----LGIYD-----ADGDGPDVDDAKVL----- 106
DB 1091 FDTVKHEAVFDKKQQTQTEGLEBPQVSSBAVVDQTTTDTVGEBAVFDVQPEKTTVEKFD 1150
QY 107 -LGLKSTSEPAVPEREAPHTPEBOQVPEAEAPONIDEAKQIQLHEMVAHEVE 165
DB 1151 DVENQQKVISERPQOEPQEAPEVPEABAKFPDSVPSVQDSQPEVPLEEVQTPQPEIQPVE 1210
QY 166 GEDLQOEDPTGEPQOEDDEFLMATVDVDRFETLE-PEVSHETHEHYNHEVTSQD-CN 223
DB 1211 SQEATFD--TYQPEQTPOE---AKFDSPEVTEQPEFSSEPTQO--HYESASDEEPN 1262
QY 224 OMEBMMSEQENP--DS-----SEPVVEDRLHHDVDTVQYV-----YEQAAYEPL 269
DB 1263 YPDEENYFPDQPSYSDIQPSFPQYVDENYDPEENYETISKSESPQEPQVQEQP- 1321
QY 270 ENEGIEITEVTAPE--DNEVE--DSQ--VIVEEVSIFPVEEQAVPPEETNRKTDPEQK 323
DB 1322 ---GEAVFEPASAFAKFDSPVESVQDSQPEPLLEEVQTOPEIQVESQPEATFDVQPEOT 1378
QY 324 AKYKKKKPKLLNKFDKTTIAELDAEKLAKKRGIEAVNAFAKLYAKYQSPPARVYKQAQ 383
DB 1379 POEA-----KFD-----SP----- 1387
QY 384 CEDDLAEKRSNVEVLAGALETQOE--VASLPVPADLLKLSKRRSDROOFLGMRGSL 441
DB 1388 -----VETIQEPQVSSPEV---VQGNPPEERKPE-----TVL 1417
QY 442 TLQRLVQLFP--NDTSLKNDLGVGYLLIGDNDNAKYVEVLSTPNDGFAVHYGEL 498
DB 1418 EEPQADEIQPEASEESBLMELLVG-----NNSYGHYEP-----DG-EWVMAGFFG 1462
QY 499 KQA--NKIASPIYLAKEGIESGDPGTDDGRFYHL-GDAMQ-RVGNKEAVKVELGHKRG 554
DB 1463 DQQKMKDA--TYWABE-----RDYVPLIGDEVYGYNNKKGEMTWIGFYDESG 1509
QY 555 HFASV---WQSLYNNGLKAQPMWTP---KETGTETLVKSLERNMKLIRDEGLAVMDK 607
DB 1510 DWLVVDQKKNRQPRIN--EAPKFEKLGNEBYGYE---DNEMW----- 1551
QY 608 AKGLFLEEDENIREKQDMSQFTIMQGRNENACKAP 645
DB 1552 -----YDEPDESGNMLVFOSEETENINEDITKDIP 1582

QY 161 ABHEVG--EDLQEDGPTGPEQEDD-----EFLMATDV-----DREFTLEPEVS 204
 Db 92 VTKDEGAEATNMDEADGKKEQTDGVSVEDVTKMKNKNSKDNNAKADDEKETKETD 151
 QY 205 HEHEHSHVHEETVSQDCNQ-----DME-----MSEQENPPSSSEVVEDRELLHD 251
 Db 152 --EADHKAKGKEIDIQEHADANGTKQANTGDIKEBGLVDEKGTW-----DEKVEKG 203
 QY 252 TDDVTQVVEEQAVVEPLENEGIEITEVAPREDNPEVSQVIVEVS-----IFPEVEQ 307
 Db 204 DEKQVYENVEGKKEDEBEKTEVEAAKAEVDESNEVEDKESSEDENNEKVSADAKE 263
 QY 308 EVPEETNRKTDPEQAKAVKVKKKPKLLNKDKTKTKAELDAEKLARKKGIENAVNAFKEI 367
 Db 264 DEKEETNDDEDEKESKSGSKKKGK-----GTSGSGKVRKKNKTEEV-----305
 QY 368 VRKYPSRPRARYKACEDDLAEKRRSNEVLRGAI--ETVQEV-----ASLPDVPADL 418
 Db 306 --KKDABPRTPF-----SDRPVTERKSVERKVALIDKDSKEPVERKGGAYIKDIPNVA 358
 QY 419 LKLSLRSDROQFL-----GHNKSGSLTLQRLVOLFPNDTSKNDLGV-GYLLIGDNPN 472
 Db 359 NKVMRKRSPETTLKLHPILFGGRGKAAQIX-----TNLFGSGFWMHDEKK 406
 QY 473 AKKVYEVSVLPND---GPAKHVYGLILKAQNKIABSIPLYKEGIE---SGDPGTDDG 525
 Db 407 AKKVKKEKKEKTKELMECDVLDIHTKATTKKSDITTKLFEPELEKPHVTDVYGD--464
 QY 526 REFVHGDMMQRYGNKAEVAKWYELGKRGHPASVWQRLYNNVGLAKQPMWPKKNGYTE 585
 Db 465 -----TVSEKEX--SSKGAKRRK-----TPKKTSPYA 490
 QY 586 LVKSLERNKMLIRDEGLAVMDKAKGFLPEDENIRK 622
 Db 491 GSSSSKRSKAKSQKSEEAATKVVYKSLAHSDDESEEEK 527
 RESULT 14
 A33430
 h-caldesmon - chicken
 N:Alternate names: caldesmon, smooth muscle, calmodulin- and actin-binding protein
 C:Species: Gallus gallus (chicken)
 C:Date: 27-Feb-1990 #sequence, revision 27-Jun-1994 #ext change 22-Jun-1999
 C:Accession: A33430; A32642; A32445; A41064; A60461; PC2003; PX0022
 R:Hayashi, K.; Kanda, K.; Kimizuka, F.; Kato, I.; Sobue, K.
 Biochem. Biophys. Res. Commun. 164, 503-511, 1989
 A:Title: Primary structure and functional expression of h-caldesmon complementary DNA.
 A:Reference number: A33430; M01D:90026426; PMID:2803315
 A:Accession: A33430
 A:Molecule type: mRNA
 A:Residues: 1-771 <HAY>
 A:Experimental source: gizzard
 A>Note: part of this sequence was confirmed by protein sequencing
 R:Brian, J.; Imai, M.; Lee, R.; Moore, P.; Cook, R.G.; Lin, W.G.
 J. Biol. Chem. 264, 13873-13879, 1989
 A:Title: Cloning and expression of a smooth muscle caldesmon.
 A:Reference number: A32642; M01D:89340480; PMID:2760048
 A:Accession: A32642
 A:Molecule type: mRNA
 A:Residues: 1-318, 334-771 <BRY>
 A:Cross-references: GB:J04968; NID:9212656; PIDN:AAA49067.1; PID:9212657
 A>Note: the authors translated the codon GAA for residue 743 as Leu
 R:Note: this alternative splice form is a high molecular weight caldesmon
 R:Hayashi, K.; Yamada, S.; Kanda, K.; Kimizuka, F.; Kato, I.; Sobue, K.
 Biochem. Biophys. Res. Commun. 161, 38-45, 1989
 A:Title: 35KDa fragment of h-caldesmon contains two consensus sequences of the tropomyo
 A:Reference number: A32445; M01D:89273666; PMID:2730665
 A:Accession: A32445
 A:Molecule type: mRNA
 A:Residues: 466-771 <HA2>
 A:Cross-references: GB:M26684; NID:9211897; PIDN:AAA48811.1; PID:9211898

R:MaK, A.S.; Carpenter, M.; Smillie, L.B.; Wang, J.H.
 J. Biol. Chem. 266, 19971-19975, 1991
 A:Title: Phosphorylation of caldesmon by p34(cdc2) kinase. Identification of phosphoryl
 A:Reference number: A41064; M01D:92041815; PMID:1939059
 A:Accession: A41064
 A:Molecule type: protein
 A:Residues: 597-600,678-696,711-721 <MAX>
 R:Wang, C.I.A.; Wang, L.W.C.; Lu, R.C.
 Biochem. Biophys. Res. Commun. 162, 746-752, 1989
 A:Title: Caldesmon has two calmodulin-binding domains.
 A:Reference number: A60461; M01D:89334885; PMID:2757638
 A:Accession: A60461
 A:Molecule type: protein
 A:Residues: 2-17, 'X', 19-38, 466-485 <MAN>
 R:Haruna, M.; Hayashi, K.; Yano, H.; Takeuchi, O.; Sobue, K.
 Biochem. Biophys. Res. Commun. 197, 145-153, 1993
 A:Title: Common structural and expressional properties of vertebrate caldesmon genes.
 A:Reference number: PC2003; M01D:94071934; PMID:8250319
 A:Accession: PC2003
 A:Molecule type: DNA
 A:Residues: 74-419 <HAR>
 R:Takagi, T.; Yazawa, M.; Ueno, T.; Suzuki, S.; Yagi, K.
 J. Biochem. 106, 778-783, 1989
 A:Title: Amino acid sequence studies on cyanoagen bromide peptides of chicken caldesmon
 A:Reference number: PX0022; M01D:90130380; PMID:2613684
 A:Accession: PX0022
 A:Molecule type: protein
 A:Residues: 462-477, 'D', 479-563, 674-762, 'A', 763-771 <TAK>
 C:Comment: This protein plays a vital role in the regulation of smooth muscle and nonmu
 C:Comment: The binding of caldesmon to F-actin is modulated by calcium and calmodulin.
 C:Comment: Two calmodulin molecules can bind to nonoverlapping domains of each caldesmo
 C:Superfamily: caldesmon
 C:Keywords: actin binding; alternative splicing; calmodulin binding; cytoskeleton; musc
 F:1-771/Product: h-caldesmon #status predicted <HAY>
 F:1-318,334-771/Product: h-caldesmon, alternative splice form #status predicted <LMAT>
 F:266-390/Region: 13-residue repeats
 F:511-582/Region: tropomyosin binding
 F:622-636/Region: tropomyosin binding
 F:597,682,717/Binding site: phosphate (Thr) (covalent) (by cdc2 kinase) #status experimenta
 F:688,711/Binding site: phosphate (Thr) (covalent) (by cdc2 kinase) #status experimenta

Query Match 4.5%; Score 181; DB 1; Length 771;
 Best Local Similarity 22.8%; Pred. No. 0.066;
 Matches 150; Conservative 87; Mismatches 248; Indels 174; Gaps 32;

QY 108 GIKERSTSE---PAVPPEE-----AEHTPEEQV-----PVEAE-----PCN1 143
 Db 178 GKKEKDESEKPKRPVPEHNVDAVERSTDEKEVETKTLAVNANTNMLEGGQSI 237
 QY 144 EDEAKEQIQLIHENVAHEVGEPLQ--QEDGPTGEPQEDDEFLMATVDVDFETLEPE 202
 Db 238 TPAADKKEKEAEKEREKLEAEKEKRLKAEKKAEEKOKAKSEKKAEREERAKAEERK 297
 QY 203 VSHEETSHSYHEEVVSQDCNQDMEEMSEQENPPSSSEV--VEDRLHDTDVTQVYE 261
 Db 298 RAEEERERAKAE--RKAERERAKAEERKAEERKAEERKAEERKAEERKAEERK 344
 QY 262 EQAVVEPLENEGIEITEVTAPEEDNPVEDSQVIVEVSIFPYVEQOEVPETNR-----315
 Db 345 ERAKAE-----EKKKAERERAKAEERKAEERKAEERKAEERKAEERKAEERK 395
 QY 316 --KTDDPEOKAVKVKKKPKLLNKDKTKTKAELDA-----AEFL-----351
 Db 396 KEKKKMAEEKKQOEFAQANTLARKOEDKAKYAEAKKESLPETLOPTSKDQVKNDKKEK 455
 QY 352 -----RKRKITE--AVNAFKELVKPKPSRPRARYGAACEDDLAEKRRSNEVLR 399
 Db 456 AKKEKSYWMDKKRGVPEOKAONGERELTPPKLKTSTENAFGSKNLK--GAANNAEGSEK 514
 QY 400 GAIEETQEVASLPDVPADLKLKLSKRSRDOQFLGMRGSLTLQRLVOLFPNDTSKND 459
 Db 515 -----EKQGEAA-----VELDEIK--KRREERRKIL-----EEBEOKKK 546

```
QY 460 LGVGYLLIGDNDNAKVEEV-----LSVTPNDGPA---KVHYGTLTKAQN-KLA 505
DB 547 QEARERKIREBERKRRKMEIERRRAEAERQKVPEDGVSEKKPKCFSPKGSLSKIE 606
QY 506 ESIPYLKSGISGDDPTDGRFYFLGDAMQVGNK-EAYKMYELGHRKG----- 554
DB 607 ERAEFLNKSQK-----SGMKPAHTAVSGKIDSRLEQYTSAVVGNKAAPAKPAPASDL 660
QY 555 -----HFASVQR-SLYVNGVIGAKQPMWTPKET-GYTELVKSLERNMKLIRDEGLAV 604
DB 661 FVPAGVGNIKSMKMGKVFSSPGGTGP---NKETAGLKVGVSSRINEMWLTKTPEG--- 714
QY 605 MDKAKGLFLPDEMLREKGDWS-OPTLMQGRRNENACKGAPKCTTLLEKFPETTGCR 662
DB 715 -NKSRA---PKPSDLR-PGDVSGKKNLMKQSEKPAASSKVTAT--GKASETNGLRQ 766
```

RESULT 15

```
S47436
flagellar antigen - Trypanosoma brucei (fragment)
C/Species: Trypanosoma brucei
C/Date: 13-Jan-1995 #sequence_revision 30-Jan-1998 #text_change 03-Nov-2000
C/Accession: S47436
R/Imboden, M.; Mueller, N.; Hemphill, A.; Mattioli, X.Y.Z.; Seebeck, T.
submitted to the EMBL Data Library, August 1994
A/Description: Repetitive proteins from the flagellar cytoskeleton of African Trypanosom
A/Reference number: S47436
A/Accession: S47436
A/Molecule type: mRNA
A/Residues: 1-411 <IMB>
A/Cross-references: EMBL:Z36281; NID:G530358; PID:G530359
A/Experimental source: strain stock TREU 1285
C/Superfamily: cytoadherence-accessory protein hmw1
```

```
Query Match 4.5%; Score 179.5; DB 2; Length 411;
Best local Similarity 25.1%; Pred. No. 0.036;
Matches 83; Conservative 53; Mismatches 128; Indels 67; Gaps 16;
```

```
QY 116 EPAVPEAEHPHTPEEQVPEA-----EPQNIEDAEKQIQSLHMYNAE-HVGEGL 169
DB 11 EPQVPAEAQPEAQPEGDIAVEALELEPEQVPAEAQPE-----SAVAPEGDIAVEAL 63
QY 170 QQEDGPTGEPOQEDDEFIMATVD--DRFETLE--PEVSHETEHSHVEETVSQDCNQDM 226
DB 64 ELEPEPQAPAEAPALPEGDIAVEALELEPEQVPAEAQPEAVAPBGDIAVEALEEL 123
QY 227 EEMNSEQENPDSSSP-VVEDRLHDDTDVTVQYEE-----QAVYEPLENEG--- 273
DB 124 EE---PQVPAEAQPEAVAPE-----GDIAVEALELEPEQVPAEAQPEAVAPBGDIA 174
QY 274 IEITEVATPEEDNVE-----DSQVIVREVSTFPVEEQQEVPPETNRKTDDEPQKAY 326
DB 175 VEALPELEPEQVPAEAQPEAQPEGDIAVE--ALELEPEQVPAEAQPEAVAPBGDIAV 232
QY 327 KKKKPKLINKEDKTIKAEIDAEKLRKRGKIEAVNAFKELVRKYPOSPPARYGKAQCED 386
DB 233 -----FALELEPEPQAPAEAPAEQPEGDI--AVEALEEL-EEPQAPAEAPAEQPEG 284
QY 387 DLAKRRSNVYLGALETYQEVASLPDVPAD 417
DB 285 DI-----AVEALELEPEPQAPAE 303
```

Search completed: May 19, 2004, 16:00:36
Job time : 34.7034 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 19, 2004, 15:51:04 ; Search time 19.2142 Seconds

(without alignments)
2054.166 Million cell updates/sec

Title: US-09-903-199-2

Perfect score: 4022

Sequence: 1 MAQRKAKSSGNSSSGSGS.....IDVWHPELTPOQRSLPAI 758

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	396.5	98.6	757	1 ASPH_HUMAN	Q12797 homo sapien
2	317.2	78.9	754	1 ASPH_BOVIN	Q28056 bos taurus
3	203	5.0	700	1 TRDN_CANIN	P82179 canis faml
4	194	4.8	4835	1 MDNI_GALIA	O81511 galidia lam
5	193	4.8	493	1 ECXI_METMA	O8P188 methanosarc
6	189.5	4.7	1616	1 P20C_MYCE	Q46429 mycoplasma
7	187	4.6	705	1 TRDN_RABIT	Q28820 oryctolagus
8	181	4.5	771	1 CALD_CHICK	P12957 gallus gall
9	178.5	4.4	728	1 TRDN_HUMAN	Q13061 homo sapien
10	176.5	4.4	845	1 NFM_RAT	P12839 rattus norv
11	175	4.4	2004	1 MYSS_HUMAN	O92794 homo sapien
12	172.5	4.3	1240	1 YNUJ_YEAST	P53935 saccharomyc
13	172	4.3	465	1 YHUE_YEAST	P38845 saccharomyc
14	172	4.3	630	1 YCF2_OENVI	P31569 cenothea v
15	169.5	4.2	592	1 LAM2_MOUSE	P21619 mus musculu
16	169.5	4.2	1233	1 YE16_YEAST	P43597 saccharomyc
17	165	4.1	650	1 NUCL_XENLA	P20397 xenopus lae
18	164	4.1	1957	1 SPOF_SCHPO	Q10411 schizosacch
19	163	4.1	411	1 FRB3_YEAST	P38911 saccharomyc
20	162.5	4.0	793	1 CALD_HUMAN	Q05682 homo sapien
21	161.5	4.0	848	1 NFM_MOUSE	P08553 mus musculu
22	161.5	4.0	1220	1 IE2P_HUMAN	O60841 homo sapien
23	161.5	4.0	1395	1 SP41_YEAST	P38904 saccharomyc
24	161	4.0	2663	1 CENE_HUMAN	Q02224 homo sapien
25	161	4.0	4910	1 MDNI_YEAST	Q12019 saccharomyc
26	160.5	4.0	721	1 YCF2_OENPI	P31568 cenothea p
27	159	4.0	3924	1 ANK2_HUMAN	Q01484 homo sapien
28	158.5	3.9	837	1 RASO_METTH	O26640 methanobact
29	158.5	3.9	1102	1 YG49_SCHPO	O60194 schizosacch
30	158	3.9	795	1 CDLI_HUMAN	P21117 homo sapien
31	157.5	3.9	780	1 CDLI_HUMAN	O94988 homo sapien
32	157.5	3.9	1357	1 KTN1_HUMAN	O66022 homo sapien
33	156.5	3.9	506	1 NPL3_HUMAN	Q99457 homo sapien

34	156.5	3.9	1658	1 YME7_YEAST	Q03661 saccharomyc
35	154.5	3.8	1828	1 MAP2_MOUSE	P20357 mus musculu
36	154	3.8	1549	1 TRHY_SHEEP	P22793 ovis aries
37	153.5	3.8	1898	1 TRHY_HUMAN	Q07283 homo sapien
38	152	3.8	434	1 YK12_YEAST	P36080 saccharomyc
39	152	3.8	671	1 CHS5_YEAST	Q12114 saccharomyc
40	152	3.8	725	1 HSB9_BRARE	O57521 brachydanio
41	152	3.8	728	1 HSB9_CHICK	P11501 gallus gall
42	152	3.8	1001	1 RPRG_MOUSE	O97085 mus musculu
43	152	3.8	1233	1 SMIA_BOVIN	O97593 bos taurus
44	152	3.8	1233	1 SMIA_HUMAN	Q14683 homo sapien
45	152	3.8	1282	1 BMS1_HUMAN	Q14692 homo sapien

ALIGNMENTS

RESULT 1
ASPH_HUMAN STANDARD; PRT; 757 AA.
ID ASPH_HUMAN
AC Q12797;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Aspartyl/asparaginyl beta-hydroxylase (EC 1.14.11.16) (Aspartate beta-hydroxylase) (ASP beta-hydroxylase) (peptide-aspartate beta-dioxygenase).
GN ASPH.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95121937; PubMed=7821814;
RA Koriath F., Gieffers C., Frey J.;
RT "Cloning and characterization of the human gene encoding aspartyl beta-hydroxylase";
RT beta-hydroxylase";
RL Gene 150:395-399(1994).
CC -!- FUNCTION: Specifically hydroxylates an Asp or Asn residue in certain epidermal growth factor-like (EGF) domains of a number of proteins.
CC -!- CATALYTIC ACTIVITY: Peptide L-aspartate + 2-oxoglutarate + O(2) = peptide 3-hydroxy-L-aspartate + succinate + CO(2).
CC -!- COFACTOR: Iron.
CC -!- SUBUNIT: Monomer (By similarity).
CC -!- SUBCELLULAR LOCATION: Type II membrane protein. Endoplasmic reticulum.
CC -!- TISSUE SPECIFICITY: Detected in all tissues tested.
CC -!- PTM: MIGHT BE PROCESSED TO THE 56 kDa (AA 274-757) OR 52 kDa (AA 315-757) FORMS IN THE LUMEN OF THE ENDOPLASMIC RETICULUM (BY SIMILARITY).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/or send an email to license@isb-sib.ch>).
CC EMBL, U03109; AAA82108.1; -.
DR PIR, I38423; I38423.
DR Genew: HGNC:757; ASPH.
DR MTM, 600582;
DR GO, GO:0005789; C:Endoplasmic reticulum membrane; TAS.
DR GO, GO:0005509; F:calcium ion binding; TAS.
DR GO, GO:0005489; F:election transporter activity; TAS.
DR GO, GO:0004597; F:peptide-aspartate beta-dioxygenase activity; TAS.
DR GO, GO:0008307; F:structural constituent of muscle; TAS.
DR GO, GO:0006936; P:muscle contraction; TAS.
DR InterPro, IPR007943; Asp-B-hydro N.
DR InterPro, IPR007803; Asp_Arg_Hydrox.


```

QY 136 VEA---EPONTIDEAKEOIQLSHMVAHEVEGEDLQOEDGPTGEOQEDDEFIMATDV 192
DB 160 AKVAHREKEKVEKEKSE-----KKAHKEKTEKKEKPEYTKMAKEREKATEKIKKEV 214
QY 193 -DDREPLEPEVS-----HEBTEH-----SYHEEVEGQDC 222
DB 215 KGGKGEKVPFAAKVKEVQKTPPKAKKEGKETAVAHGEKQDOYAFRWMIDMVFVIGDL 274
QY 223 NQDHEEM-----SEQENPDSEPVVEDRLHHTDDVYQVE 261
DB 275 RPQGSFALPPLPTVQASRPPTPASPTLEGEKEEKKKAEKKTSETKKEKEDVYKXSDK 334
QY 262 EQAV-----YEPLENEG-IEITEVTAPREDNPVEDSQVIVEVSIFFVEEOG----- 308
DB 335 DTAIDVEKEPEKAPETKQGTIKVVAQAAKKDEKEDSKTKTPVEEHFPGKQOEKKEK 394
QY 309 -VPEETNRKTD--PECK-----AKVKKKP-KLINKFDKIKAEIDAAE 349
DB 395 YVEPAKSKKSHSAPSEKQVAKTERAKEETSAASTKAVGKKEKTKTVEQEI----- 450
QY 350 KLRKRGKI-----EAVNAFKELVRYQSPPARVYK-----AQCEDD 387
DB 451 RKESGKSTASKKEPEIKDEKMPKADKEVKEPQSQVKEEKSQVKEAKPEQD 510
QY 388 LAEKRS-----NEVIRGAIE-----TYQEVASLPVPADLKLKLS 422
DB 511 IAKPEKTVSHGKPEEKVVKQVAKTEKKAIEKTVKPKAKAEHDEKKS-PITKTDKPEPT 569
QY 423 LK-----RRSDRQOFLGHRGSLTLQRLVQDFPNDTSLKNDLGVYL 465
DB 570 SKETPEVYESGKKKIEKSEKSEKAEKMKLKEKVENSRKSLQSHNTAKEXPARVSRE 629
QY 466 LIGDNDNAKKVYEVLVY-----TPNDGFAKVH-----YGFILKAQNKIAESIPYLK 512
DB 630 DLEVDASAKKXKEKEDVSTKQKQSPISFCQVYLDGNGYGFQFPV-----TPAYR 682
QY 513 EGIESGDGCT 522
DB 683 PGESSGQPS 692

```

RESULT 4

MDN1_GIALA STANDARD; PRT: 4835 AA.

AC Q81ST1,
28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DI 10-OCT-2003 (Rel. 42, Last annotation update)
DE Midasin (WIDAS-containing protein).
GN MDN1.
OS Giardia lamblia (Giardia intestinalis).
OC Eukaryota; Diplomonadida; Hexamitidae; Giardiae; Giardia.
OX NCBI_TaxID=5741;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20389616; Pubmed=10930750;
RA McArthur A.G., Morrison H.G., Nixon J.E., Passamaneck N.O., Kim U.,
Hinkle G., Crocker M.K., Holder M.E., Fair R., Reich C.I., Olsen G.E.,
Aley S.B., Adam R.D., Gillin F.D., Sogin M.L.;
RT "The Giardia genome project database."
RU FEMS Microbiol. Lett. 189:271-273(2000).
RN [2]
RP IDENTIFICATION, GENE NAME, AND SIMILARITY WITH OTHER FAMILY MEMBERS.
RX Pubmed=12102729;
RA Garbarino J.E., Gibbons I.R.;
RT "Expression and genomic analysis of midasin, a novel and highly
conserved AAA protein distantly related to dynein."
RU BMC Genomics 3:18-18(2002).
CC -!- FUNCTION: May function as a nuclear chaperone and be involved in
the assembly/disassembly of macromolecular complexes in the
nucleus.
CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC -!- SIMILARITY: Contains 1 VFMA domain.

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements/
CC or send an email to license@ebi.ac.uk).
CC -----
DR EMBL; AF494287; AAM12656.1; -.
DR InterPro; IPR003593; AAA ATPase.
DR InterPro; IPR003959; AAA ATPase_centre.
DR InterPro; IPR002035; VFMA_A.
DR Pfam; PF00004; AAA; 2.
DR SMART; SM00382; AAA; 5.
DR PROSITE; PS50234; VFMA; 1.
KW Chaperone; ATP-binding; Repeat; Nuclear protein.
FT NP_BIND 356 363 ATP (POTENTIAL).
FT NP_BIND 814 821 ATP (POTENTIAL).
FT NP_BIND 1127 1134 ATP (POTENTIAL).
FT NP_BIND 1513 1520 ATP (POTENTIAL).
FT NP_BIND 1839 1846 ATP (POTENTIAL).
FT NP_BIND 3277 3284 ATP (POTENTIAL).
FT DOMAIN 110 113 POLY-LEU.
FT DOMAIN 3004 3007 POLY-ASN.
FT DOMAIN 4156 4161 POLY-ASN.
FT DOMAIN 4629 4818 VFMA.
SQ SEQUENCE 4835 AA; 539726 MW; 3A9E12417DB04A50 CRC64;

Query Match 4.88; Score 194; DB 1; Length 4835;
Best Local Similarity 23.1%; Pred. No. 0.084;
Matches 119; Conservative 74; Mismatches 203; Indels 120; Gaps 25;

QY 80 YREYVGLKGIYD--ADGSDPVDVDAKVLGLKERSTSEPAVPEEAEPHTEPEQVVE 137
DB 4188 HEBGADATGSTDDEAQEDVDYNDLDD-KNLGG-----QSDLSVPADDEDEVNE--LE 4238
QY 138 AEPONTIDEAKEOIQLSHMVAHEVEGEDLQOEDGPTGEOQEDDEFIMATDVDFRE 197
DB 4239 EEOQOMSDLNPD-----QDACAIEEDDDRLDPSD-----ENAEHDEHEAPVIDDN-E 4288
QY 198 TLPEVSHETSHSYHEVEIVSQDQCNQDMEEMSECE-----NPDSEPVVEDER- 247
DB 4289 ASDQSTYNDNRDADANISAQQAINDDEEMOKTLYDQENTDTSNPDAAVEGTNDQK 4348
QY 248 LHHDTDDVYQVEEQAVVEPLEN---EGEITEVTAPREDNPVEDSQVI-VEEVSIFPV 303
DB 4349 THEHNDQFRQENIEDQWEASTENSGEGASADLKEGNDPMSLEEFQRIWKERTNIHDR 4408
QY 304 E-----EQQVPEETNRKTDDEBQKA-----KYKKKKPKLLNFKDKIKAEIDAAE 349
DB 4409 ESEKDEAEAPQDMDLQSNKTVKEPDSKSGDGLGTLSTSHRNLTNQ-----EFDPVN 4461
QY 350 KLRKRGKIEEAVNAFKELVRYQSPPARVYKXQCEPDAAE-----KRR 393
DB 4462 EER--NVEH--NSCCTSGSHDRPAEHLNPEISDGESESSASDKQOQAVLSHAKES 4516
QY 394 SNEVIRGAIETVQVEVA-SLPD-----VPADL-----LKSLSKRSRQOFLGHRG 438
DB 4517 SKDLINNEGAVYQGLAVSLASEETKRAPEVVAASARGNHLLDLIKQTSAAAE----- 4570
QY 439 SLTLQRLVQVFPNDIS-LKNDLGVYL-----IGNDNAKKVYE 478
DB 4571 SLAEKRLII-LPEVTDLKDQDFGTGKLLNRIIPFASFPQDKIWLRTKFSKRYQ 4629
QY 479 EVLSVTNPDGFAKVHGFILKAQNKIAESIPYLKRG 514
DB 4630 VILAVDDSSMAPI-AKVALQAITLLFNACKFLEVG 4664

```

RESULT 5

ECX1_METMA

ID ECK1_METWA STANDARD; PRT; 493 AA.
AC Q8PT8;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Probable exosome complex exonuclease 1 (EC 3.1.13.-).
GN MM2623.
OS Methanosarcina mazei (Methanosarcina frisia).
OC Archaea; Euryarchaeota; Methanocicrobia; Methanosarcinales;
OC Methanosarcinaceae; Methanosarcina.
OX NCBI_TaxID=2209;
[1]
SEQUENCE FROM N.A.
RP STAIN=Geel / G01 / ATCC BAA-199 / DSM 3647 / OCM 88;
RC MEDLINE=22120827; PubMed=12125824;
RA Deppeleier U., Johann A., Hartesch T., Merkl R., Schmitz R.A.,
Martinez-Arias R., Hense A., Wierzer A., Baumer S., Jacob C.,
Brueggemann H., Uenard T., Christmann A., Boemecke M., Steckel S.,
Rahatsharyya A., Lykdis A., Overbeek R., Klenk H.-P., Gunsalus R.P.,
Fitz H.-J., Gottschalk G.;
RA "The genome of Methanosarcina mazei: evidence for lateral gene
transfer between Bacteria and Archaea";
RT J. Mol. Microbiol. Biotechnol. 4:453-461(2002).
CC -1- FUNCTION: Probably involved in the 3'->5' degradation of a variety
of RNA species (Potential).
CC -1- SUBUNIT: Component of the archaeal exosome multienzyme
ribonuclease complex (Potential).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC -1- SIMILARITY: Belongs to the RNase PH family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

DR EMBL; A013507; AAM32319.1; -.
DR HAMAP; MF 00591; -; 1.
DR InterPro; IPR001247; 3 EXORNAse.
DR Pfam; PF01138; RNase_PH; 1.
DR Pfam; PF03725; RNase_PH_C; 1.
KW Exosome; Hydrolase; Nuclease; Exonuclease; Complete proteome.
FT DOMAIN 1 254 PROBABLE EXOSOME COMPLEX EXONUCLEASE 1.
FT DOMAIN 255 493 UNKNOWN.
SQ SEQUENCE 493 AA; 55248 MM; 979A757BF80C930 CRC64;

Query Match 4.8%; Score 193; DB 1; Length 493;
Best Local Similarity 26.3%; Pred. No. 0.0056;
Matches 97; Conservative 45; Mismatches 129; Indels 98; Gaps 17;

QY 61 IALLGWTSTVAVVWFDLVDEEVGKGIYDADGDGDFV---DDAKVL----- 107
DB 153 LPMKGLITSCA---FGYDKIVLDLNKEDNTEADFPVAMTODGEITILIQMDGNLTDP 209
QY 108 -----GLKERTSEPAV-----PPEEAEPTTE-----EOVVEAPQ 141
DB 210 EIKQGLGVKKGCKELILEIQAVLRKKFPEVVEVSETPAKKAKEKVLSPVAIVE 269
QY 142 NIEDEAKQIOLSLHENVHAEVGEEDLQOEDGTGPQOD--DEFLMAT--DVDRFE 197
DB 270 ETPEAESEP-----EVEISREVAEAILASVTPDFEDDELEEELEESSEEDLETEE 323
QY 198 TLEPEVHEETESHVVEETVSDQDMEMMSSEGNPSSSEVVDE----- 246
DB 324 EFEEBALEBEAEPEEDLEEDLGELEBELEEELEEEBEAELEAELEASLECAP 383
QY 247 -----RLHNDTDVTVQVYE-----EQAVPELENEGIEITVAPEDNPEVD 290
DB 384 ELKFEDEIEARL--EKEDASIEAEIEEIEPEAEKTEGLEBEA-EIETAASEEN--IEA 439
QY 291 SQVIVEVSIFFVVEQOQVPEPTNRKTDDPEQAKVKKKKPKLNRDKITIKALDAEK 350

DB 440 EAAEEAEAE--PEVEAEISTEAEAEAEAEAE-----KSGPW-----KVXDPSEAG-- 486
QY 351 LRRGKIEE 359
DB 487 --TRGEKDE 493

RESULT 6
P200 MYCGE
ID P200 MYCGE STANDARD; PRT; 1616 AA.
AC Q49429; Q49259; Q49296; Q49352; Q49353;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Protein P200.
GN MG386.
OS Mycoplasma genitalium.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2097;
[1]
SEQUENCE FROM N.A.
RP SRRAIN=ATCC 33530 / G-37;
RC MEDLINE=96026346; PubMed=7569993;
RA Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,
Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,
Fitchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhmann J.L.,
Nuyen D.T., Uterback T.R., Saudek D.M., Phillips C.A., Merrick J.M.,
Tombl J.-F., Dougherty B.A., Bort K.F., Hu P.-C., Lucier T.S.,
Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;
RT "The minimal gene complement of Mycoplasma genitalium";
RL Science 270:397-403(1995).
[2]
SEQUENCE OF 256-427; 432-543 AND 1083-1140 FROM N.A.
RP SRRAIN=ATCC 33530 / G-37;
RC MEDLINE=94075230; PubMed=8253680;
RA Peterson S.N., Hu P.-C., Bort K.F., Hutchison C.A. III,
RT A survey of the Mycoplasma genitalium genome by using random
sequencing";
RT J. Bacteriol. 175:7918-7930(1993).
CC -1- FUNCTION: COULD BE AN ACCESSORY STRUCTURAL COMPONENT IN
CYTADHERENCE (BY SIMILARITY).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

DR EMBL; U03920; AAC71613.1; -.
DR EMBL; U02245; AAA03400.1; -.
DR EMBL; U02245; -; NOT ANNOTATED_CDS.
DR EMBL; U02175; AAD12458.1; -.
DR EMBL; U02126; AAD12402.1; -.
DR PIR; G64242; G64242.
DR TIGR; MG386; -.
KW Cytoplasmic; Structural protein; Repeat; Complete proteome.
FT DOMAIN 1205 1389 2 X 32 AA REPEAT.
FT REPEAT 1205 1236 1-1.
FT REPEAT 1358 1389 1-2.
FT DOMAIN 891 1389 2 X 26 AA REPEAT.
FT REPEAT 1161 1186 2-1.
FT REPEAT 1310 1339 2-2.
FT REPEAT 1310 1339 P -> S (IN REP. 2).
FT CONFLICT 256 255 P -> P (IN REP. 2).
FT CONFLICT 304 304 S -> P (IN REP. 2).
SQ SEQUENCE 1616 AA; 185678 MM; 6AF76A13AC49E4FF CRC64;

Query Match 4.7%; Score 189.5; DB 1; Length 1616;
Best Local Similarity 22.3%; Pred. No. 0.037;
Matches 142; Conservative 78; Mismatches 205; Indels 213; Gaps 35;

```

QY 75 FDLYDYEVYGLK-----LGTYD-----ADGSDGFVDYDAKVL----- 106
DB 1091 FDYVGHAEVAFKQNTQOTEGLEBPQVSSAEAVDDQTTTDTTVEBPVAVQCEKTEVYFD 1150
QY 107 -LGKERSTSEPAVPEAEAPHTPEBQVPEAEQONIEDAKEDICQLHEMNAHEVE 165
DB 1151 DVENQOKVISEQVQVQGEAVFESAFAKDPSPVESQDSQDFEVEVQTPETQVPE 1210
QY 166 GEDLQOEDGPTGEPQOEDDEFLMATVDVDRFETLE-PEVSHETESHVYEETVSOD-CN 223
DB 1211 SOEPATFD--TVQEPQTPQE-----AKFDSPVETVQEPFSEEPQO--HVSEKSPDEPN 1262
QY 224 QDMEMMEBQENP--DS-----SEPVYEBRLHDTVDVTVQY-----YEQAVYEP 269
DB 1263 YDFEBPNYDFQPSYDSDLPSEBPQYDVDEPNYDFDEPNYIESKSPSEBPQVPEQBP- 1321
QY 270 ENEGIEITEVYAPPE-DNPVE---DSQ--VVEEVSIFPEBQOQVPEPTKTDDEPQK 323
DB 1322 ---GEAVFESAFAKDPSPVESQDSQEPFLIEVQTPETQVPEBQEPATFTVQEPQ 1378
QY 324 AKVKKKKPKLTKEDKTIKAEILDAAEKLRKQKTEBAVNAFKEIVRKYPQSPRARVKAQ 383
DB 1379 PQEA-----KFD-----SP----- 1387
QY 384 CEDDLAEKRSNEVIRGAIETVQE--VASLPDVPADLIKLSLKRSDRQOFLGMRGSL 441
DB 1388 -----VETIOEQVSSPEEV---VQPNFEEKPE-----TVL 1417
QY 442 TLQRLVLPF---NDTSIKNLIQGYLLIGNDNAKRYEVLSTPNDGSAKHYGIL 498
DB 1418 EEPQDEIQLPASEESLDMELLVG-----NNSYGYEP-----DG-EMVWAGPFG 1462
QY 499 KAO--NKIAESIPYLKESIGSDPGTDGRFYFH-GDAMQ-RVGNKEAYKMYELGHRG 554
DB 1463 DDQKKNKA-TVKNARE-----RDYLPILGEVYGRNKNKEWMTWYFYESG 1509
QY 555 HFASY---WQSLINNGIKAKQPMWTP---KETGYELVKSLEPNKILIDEGLAWDX 607
DB 1510 DMVAVDEQWKNRQPRIN--EAPKFEKXIGNEEYGYE--DNEWMW----- 1551
QY 608 AKGLFPEDENLREKGMDSOFTMOQGRNEMACKGAP 645
DB 1552 -----YDGEFDEGNMLVFOSEETENLMDITKQIP 1582

```

```

RL FEBS Lett. 348:17-20 (1994).
RN [3]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE (ISOFORMS CARDIAC).
RC TISSUE=Heart muscle;
RX MEDLINE=6132942; PubMed=8550602;
RA Guo W., Jorgensen A.O., Jones L.R., Campbell K.P.;
RT "Biochemical characterization and molecular cloning of cardiac
triadin.";
RL J. Biol. Chem. 271:458-465 (1996).
RN [4]
RP CARBOHYDRATE-LINKAGE SITES, AND INTERCHAIN DISULFIDE BONDS.
RX MEDLINE=5606664; PubMed=7578102;
RA Fan H., Brandt N.R., Caswell A.H.;
RT "Disulfide bonds, N-glycosylation and transmembrane topology of
skeletal muscle triadin.";
RL Biochemistry 34:14902-14908 (1995).
CC -1- FUNCTION: May be involved in anchoring calsequestrin to the
junctional sarcoplasmic reticulum and allowing its functional
coupling with the ryanodine receptor.
CC -1- SUBUNIT: Homooligomer of variable subunit number; disulfide-
linked.
CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Sarcoplasmic
reticulum.
CC -1- ALTERNATIVE PRODUCTS:
Event=Alternative splicing; Named isoforms=6;
Comment=Additional isoforms seem to exist;
Name=Skeletal 1; Synonyms=ST1;
IsoId=Q28820-1; Sequence=Displayed;
Name=Cardiac 1; Synonyms=CT1;
IsoId=Q28820-2; Sequence=VSP_004458, VSP_004460;
Name=Cardiac 2; Synonyms=CT2;
IsoId=Q28820-3; Sequence=VSP_004459, VSP_004461;
Name=Cardiac 3; Synonyms=CT3;
IsoId=Q28820-4; Sequence=VSP_004462;
Name=Skeletal 2; Synonyms=ST2;
IsoId=Q28820-5; Sequence=VSP_004462, VSP_004463, VSP_004464,
VSP_004465;
Name=Skeletal 3; Synonyms=ST3;
IsoId=Q28820-6; Sequence=VSP_004464, VSP_004465;
CC -1- TISSUE SPECIFICITY: Skeletal and cardiac muscle.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U31540; AAC48496.1; -
DR EMBL; L10065; AAA31488.1; -
DR EMBL; U31555; AAC48497.1; -
DR EMBL; U34201; AAC48498.1; -
DR PIR; A45990; A45990.
KM Transmembrane, Sarcoplasmic reticulum; Glycoprotein;
KM Alternative splicing.
FT INTR MET 0
FT DOMAIN 1 46
FT TRANSMEM 47 67
FT DOMAIN 68 705
FT CARBOHYD 74 74
FT CARBOHYD 624 624
FT VARSPLIC 264 285
FT VARSPLIC 264 307
FT VARSPLIC 286 705
FT VARSPLIC 308 705

```


RT "identification of six phosphorylation sites in the COOH-terminal
 RT tail region of the rat neurofilament protein M.";
 RL J. Biol. Chem. 267:4467-4471 (1992).
 RN [4]
 RP CARBOHYDRATE-LINKAGE SITES.
 RX MEDLINE=93346421; PubMed=8344946;
 RA Dong D.L.-Y., Xu Z.-S., Chevrier M.R., Cotter R.J., Cleveland D.W.,
 RA Hart G.W.;
 RT "Glycosylation of mammalian neurofilaments. Localization of multiple
 RT O-linked N-acetylglucosamine moieties on neurofilament polypeptides
 RT L and M.";
 RL J. Biol. Chem. 268:16679-16687 (1993).
 CC -1- FUNCTION: Neurofilaments usually contain three intermediate
 CC filament proteins: L, M, and H which are involved in the
 CC maintenance of neuronal caliber.
 CC -1- PFM: There are a number of repeats of the tripeptide K-S-P, NFM is
 CC phosphorylated on a number of the serines in this motif. It is
 CC thought that phosphorylation of NFM results in the formation of
 CC interfibrillar cross bridges that are important in the maintenance
 CC of axonal caliber.
 CC -1- PFM: Phosphorylation seems to play a major role in the functioning
 CC of the larger neurofilament polypeptides (NF-M and NF-H), the
 CC levels of phosphorylation being altered developmentally and
 CC coincident with a change in the neurofilament function.
 CC -1- SIMILARITY: Belongs to the intermediate filament family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M18628; AAA41696.1; -;
 DR EMBL; Z12152; CAA78136.1; -;
 DR PIR; A45669; A45669.
 DR GLYCOSULEDB; P12839; -;
 DR InterPro; IPR006821; Filament_head.
 DR InterPro; IPR001664; IF.
 DR InterPro; IPR002957; Keratin_I.
 DR Pfam; PF00038; Filament; 1.
 DR Pfam; PF04732; Filament_head; 1.
 DR PRINTS; PR01248; TYPEKERATIN.
 DR PROSITE; PS00226; IF; 1.
 KM Intermediate filament; Coiled coil; Neurone; Phosphorylation;
 KM Glycoprotein.
 FT INIT MET 0
 FT DOMAIN 1 103
 FT 103 410
 FT 411 844
 FT 103 134
 FT 135 147
 FT 148 246
 FT 247 263
 FT 264 285
 FT 286 289
 FT 290 410
 FT 410 47
 FT 47 430
 FT 430 502
 FT 502 506
 FT 506 536
 FT 536 603
 FT 603 608
 FT 608 666
 FT 666 17
 FT 17 21
 FT 21 204
 FT 204 500
 FT 500 845 AA; 95660 MW; 316C41655B1197D CRC64;

Query Match 4.4%; Score 176.5; DB 1; Length 845;
 Best Local Similarity 23.3%; Pred. No. 0.074;
 Matches 90; Conservative 59; Mismatches 137; Indels 101; Gaps 15;
 QY 81 EEVLGKGIYDADGDFVDVDAKVLGLKRSITSEPAVPEEAPPHTEPEQVPEAE 140
 DB 453 EEIEETKVEDEKS-----EMEDALTVIAEELIAASAK-----EKEKEAEKEKEPEVEKSP 503
 QY 141 QNIDKAEKQIQLHHEVNAHEVHGDLQGEDPTEPQODEPQODEPFLMATVDREFELLE 200
 DB 504 VK-SPEAKER-----BEGKEKEEBQ--EEEBEDGVKSDDAEBGGS--E 545
 QY 201 PEVSEETERSHYVEETVSQDNQDMSEMSQENPDSPSEPVDERLHDTDTDTQVY 260
 DB 546 KEGSEKDE-----GEQEBEGTANBGEBAKAEKKTBEKV 584
 QY 261 EEQAVTEPLENEGIE-----ITETVAPPDQNVESQVVEEVSITPVEEQQVP 310
 DB 585 EEMAIKEIKVEKEKAKSPVPSVPEVEKPEAKGKQKEEK-----VEKKEVA 639
 QY 311 PET--NRKTDDPQOKA-----VKKKKPKLTKPKTKAEIDAAEKLRK---- 353
 DB 640 KESPEKEKVEKEKRPDVPDKKASPVKEKAVEMITTKSVKSLKXTKEKPPQO 699
 QY 354 --RGKIEAVNAFKEELVRKYPQSPRAR-----YKACQEDDLAKRRSNEVLRGAI 402
 DB 700 EKVEKEKEEBEGSEEGEDKSPQSKKEDIALINGEVEKEEBEGTQEKSGSEKGV 759
 QY 403 ETYQEVASLDVPADLKLKLRSDR 429
 DB 760 TNGIDVS-----PAE-----EKGEDR 776
 RESULT 11
 MY33_HUMAN STANDARD; PRT; 2004 AA.
 ID MY33_HUMAN
 AC Q92794;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE MY33 histone acetyltransferase 3 (Runt-related transcription factor
 DE binding protein 2) (Monocytic leukemia zinc finger protein) (Zinc
 DE finger protein 220) (MOZ2 OR MOZ).
 GN MY33 OR RUNXBP2 OR ZNF220 OR MOZ.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=96376968; PubMed=8782817;
 RA Borrows J., Stanton V.P., Jr., Andresen J.M., Becher R., Behm F.G.,
 RA Chaganti R.S.K., Clavin C.I., Distchech C., Dube I., Fritchaut A.M.,
 RA Horsman D., Mielman F., Volinia S., Watson A.E., Housman D.E.;
 RT "The translocation t(8;16)(p11;p13) of acute myeloid leukaemia fuses
 RT a putative acetyltransferase to the CREB-binding protein.";
 RL Nat. Genet. 14:33-41 (1996).
 CC -1- FUNCTION: May represent a chromatin-associated acetyltransferase.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- DISEASE: Involved in acute myeloid leukemias through a chromosomal
 CC translocation t(8;16)(p11;p13) involving MY33 and CREBBP.
 CC -1- SIMILARITY: Contains 2 PHD-type zinc fingers.
 CC -1- SIMILARITY: Belongs to the MY33 (SAS/MOZ) family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; U47742; AAC50662.1; -
 DR Genew; HENC:13013; MYST3.
 DR MIM; 601408; -
 DR GO; GO:0006323; P:DNA packaging; TAS.
 DR InterPro; IPR005818; Histone_H1/H5.
 DR InterPro; IPR002717; MOZ_SAS.
 DR InterPro; IPR01965; Znf_PHD.
 DR Pfam; PF01853; MOZ_SAS; 1.
 DR Pfam; PF00628; PHD; 2.
 DR SMART; SM00526; H15; 1.
 DR SMART; SM00249; PHD; 2.
 DR PROSITE; PS01359; ZF_PHD_1; 1.
 DR PROSITE; PS0016; ZF_PHD_2; 2.
 DR Piro- oncogene; Chromosomal translocation; Zinc-finger; Repeat;
 KW Nuclear protein.
 FT ZN_FING 206 265 PHD-TYPE 1.
 FT ZN_FING 259 313 PHD-TYPE 2.
 FT DOMAIN 371 379 POLY-SER.
 FT ZN_FING 538 560 POLY-SER.
 FT DOMAIN 788 801 C2HC-TYPE.
 FT DOMAIN 989 995 POLY-GLU.
 FT DOMAIN 1019 1026 POLY-ARG.
 FT DOMAIN 1069 1078 POLY-GLU.
 FT DOMAIN 1147 1150 POLY-LYS.
 FT DOMAIN 1221 1242 GLU-RICH.
 FT DOMAIN 1267 1302 GLU-RICH.
 FT DOMAIN 1411 1414 POLY-GLU.
 FT DOMAIN 1593 1597 POLY-SER.
 FT DOMAIN 1643 1704 GLN/PRO-RICH.
 FT DOMAIN 1897 1977 MET-RICH.
 FT SITE 1546 1547 BREAKPOINT FOR TRANSLOCATION TO FORM MOZ-CBP.
 SQ SEQUENCE 2004 AA; 225054 MW; 9FEBAC3792854BA CRC64;
 Query Match 4.4%; Score 175; DB 1; Length 2004;
 Best Local Similarity 22.7%; Pred. No. 0.26; Mismatches 113; Indels 88; Gaps 13;
 Matches 73; Conservative 48; Mismatches 113; Indels 88; Gaps 13;
 QY 122 EEAPEHPTEPEEQVEAEAPQNIIDEAKEQISLHEWVAHEVGEEDLQEDGPTGEPQ 181
 DB 1205 QESSELYEPKEDM-----LPREKKEE-BEMQAEAEAEAGEEDAAAEVAAAPAC 1256
 QY 182 EDDEFLMATVDVDFETLEPEVSHETESHVHEVTSQDCNQDEEMMSQENPDSEEP 241
 DB 1257 SSN-----SPETETKEPEVEEEEEKPRVSEEQROSEEQLEPEPEEEDAAET 1308
 QY 242 VVEDERLHNDTD-----VTVQVVEQAVVEPELNE-GI----- 274
 DB 1309 AQNDU---HDADDDDGHLSTKKKELEEOPTREDVKEEPGVQESFLDANMOKSREKIK 1365
 QY 275 -EITEVTAPEPDNFEVDSEQVIVEEVSIFPVBEQOEVPETNRKTDDEQKAKVKKKKPKL 1365
 DB 1366 KEELHELDS-EEEQSHDSVSSEQA-----GSEDDHEDSHTEKELIEL 1409
 QY 334 LNKEDKTTKAELEDAEKLRKRGKTEBAVNAFKELVRKPPSPARVYGAQCEDDLAEKRR 393
 DB 1410 KEE-EEIPHSLELD-----ETVQAVQSLOESSRHGAY--QDCEETLA----- 1451
 QY 394 SNEVLRGALIEYQEVASLPVPE 415
 DB 1452 -----ACQTLQSYTQADEDF 1466
 RESULT 12
 YNJI_YEAST STANDARD; PRT; 1240 AA.
 AC P53935
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE Hypothetical 141.5 kDa protein in YP53-RHO2 intergenic region.
 GN YN1091W OR N2231.
 OS Saccharomyces cerevisiae (Baker's yeast).

OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 CC NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=5288C / FY1679;
 RA MEDLINE=96367601; PubMed=8771715;
 RX Garcia-Cantalejo J M, Boskovic J, Jimenez A;
 RT "Sequence analysis of a 14.2 kb fragment of Saccharomyces cerevisiae
 chromosome XIV that includes the YPC53, TRN146 and gsr m2 genes and
 four new open reading frames."
 RT Yeast 12:599-608 (1996).
 RL -1- SIMILARITY: TO S.POMBE SPAC29E6.10C.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements>
 CC or send an email to license@ebi.ac.uk).

DR EMBL; X65811; CAAS9826.1; -
 DR EMBL; Z71367; CAAS9567.1; -
 DR PIR; S52734; S52734.
 DR GerMOnline; 143097; -
 DR SGD; S0005035; YN1091W.
 DR GO; GO:0009651; P:salinity response; IMP.
 KW Hypothetical protein.
 FT DOMAIN 756 761 POLY-GLU.
 SQ SEQUENCE 1240 AA; 141513 MW; 3FE9D26582D5778 CRC64;
 Query Match 4.3%; Score 172.5; DB 1; Length 1240;
 Best Local Similarity 23.9%; Pred. No. 0.19; Mismatches 175; Indels 87; Gaps 21;
 Matches 103; Conservative 66; Mismatches 175; Indels 87; Gaps 21;
 QY 144 EDEAKEQISLHEWVAHEV-GEEDLQEDGPTGEBQEDDEFLMATVDVDFETLEPE 202
 DB 460 EDEDEEDYD-----YSEVAEDSEVESEYGEIYAEKPEHDE-----KSNIGRETLILS 508
 QY 203 VSHE-----ETESHVHEVTSQDCNQDEEMMSQENP-----DSEEP 241
 DB 509 YDHDKRKQNPNNHHYSTHSEED-ELSEELYSIDILPHDPKHFRRDDILDGDEDEP 567
 QY 242 VVEDERLHNDTDVTVQ--VYEQAVYEPLENBEGITTE-----VTAPEDNPEVDQV- 293
 DB 568 EBEEDENGDEBD- TVYSGIDETDRLEBGRKLIQIATTKLQSLMASYNHKKQADNNRIK 626
 QY 294 IVEEVSIFPVBEQOEVPETNRKTDDEQKAKVKKKKPKLNFKDK-----TIAEL 345
 DB 627 LLOE-----LEEEKRKREKEEKKQKREKREKRLQQLAKEBEKRRREBEKELTKEL 681
 QY 346 DAEEKLRK--RGKIEAVNAFKELVRKPPSPARVYGAQCEDDLAEKRR--SNEVLRG 400
 DB 682 EBEEMRRREKQKKEVEA-----KKKQDEBRGRLEBQRRREBEKQKQKEBELRK 734
 QY 401 AIEYQVAVSLPVPADILKLSLRSDROFLGHEGSLITLQRL--VQLFPNDTSLK- 457
 DB 735 REBEKGRIRQKQLEQKQKE--KEEERGRILAE--DALRKQKLMBEQSAILLSAKP 790
 QY 458 -NDLGVGYLIGDNDNAKKVYEEVLSVPNDGFAKVHYGLAKQNKIABSIPLAKGIE 516
 DB 791 FTEGVGNPVYSQSHPMVTNYQEDNSCINDE-----ILKWNVSVAASKSPVPTGFN 842
 QY 517 SGD---PGTD 524
 DB 843 VHDLLPSTNN 853
 RESULT 13
 YHUG_YEAST STANDARD; PRT; 465 AA.
 AC P38845;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE Hypothetical 141.5 kDa protein in YP53-RHO2 intergenic region.
 GN YN1091W OR N2231.
 OS Saccharomyces cerevisiae (Baker's yeast).

```

DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hypothetical 51.1 kDa protein in DCD1-MRPL6 intergenic region.
GN YHR146W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycaceae.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=5288c / AB972;
RX MEDLINE=94378003; PubMed=8091229;
RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Dover J.,
RA Du Z., Favello A., Fulton L., Gatungu S., Geisel C., Kirsten J.,
RA Kueba T., Hillier L.W., Jier M., Johnston L., Langston Y.,
RA Latreille P., Louis E.J., Macri C., Mardis E., Meneses S., Mueier L.,
RA Nham M., Rifkin L., Riles L., St Peter H., Trevasaks E., Vaughan K.,
RA Vignati D., Wilcox L., Wohlman P., Waterston R., Wilson R.,
RA Vaudin M.;
RT "Complete nucleotide sequence of Saccharomyces cerevisiae chromosome
RT VIII.", 265:2077-2082(1994).
RL Science 265:2077-2082(1994).
CC -I- SIMILARITY: TO YEAST YML173C.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U10397; AAB68962.1; -
DR PIR: S46759; S46759.
DR GenOnline: 139464; -
DR SGD: S0001189; CRP1.
DR GO: GO:0005634; C:nucleus; IC.
DR GO: GO:0003677; P:DNA binding; IDA.
KM Hypothetical protein.
SQ SEQUENCE 465 AA; 5115 MW; 30880758F37991C7 CRC64;

Query Match 4.3%; Score 172; DB 1; Length 465;
Best Local Similarity 22.6%; Pred. No. 0.06;
Matches 90; Conservative 52; Mismatches 132; Indels 124; Gaps 16;

QY 76 DIVDYBEVLGKLGTYDADGDFDVDAKVLIG----- 108
DB 92 DIVERQEVAGASRIPEAG-----LLCGKPPRSAGPSTSNRKKNNKKRRSK 141
QY 109 LKERSTSEPAVPEEAHPHTEPE-----ROYVEAPONIEDEAKEQIQS 153
DB 142 LKKKSTKNNKKSNEJLDNDEEDGVYGTTEDEVYTSREBTL-APFNVSKEAGNPHI 200
QY 154 LHNEMVHAHVEGEDLQOEDGPTGEP-----QOEDDEFMATVDVDRFETLEPEVSH 206
DB 201 LPTD-----QSADTTQNGILGGPGLVNPGEIKEFTEIRVDAR-----F 243
QY 207 ETEHSYHVEYVSQDCNODMEEMSGEN--PDSEPYVEDERLHHTDDVYTYOYEEQA 264
DB 244 LNERLNKKKEEVEPEVAGPIVSESVTEKSPALFQADDPVETKEVANHANOELTPQV--EA 300
QY 245 VVEPLENGIEITETVAPEDNPVEDSOVIVEVS-IPFVEQOQVPPETNRKTKDDPQK 323
DB 301 V-TPINE-----PEPLTPPEAQISIPESKVEPEVSGSD----- 334
QY 334 AKVKKKKKPLNPKDKITKALIDAAKRLRKRGKIEBAVNAFELVKKYPOSP-----RAR 378
DB 335 -----SKLVEKRST-EGVLDSGSKYENKAKKQDEVPFTDPIYNAKPKPLTDEQNAH 386
QY 379 YGKAQCEDDLAKRKSNEVLRGALITTYQEVASLPVPA 416
DB 387 GRKSPAVSEERKKKKQK--KGSKEVKRSESTSEKPKS 422

```

```

RESULT 14
ID YCF2_OENV1 STANDARD; PRT; 630 AA.
AC P31569;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DE Hypothetical protein ycf2 (ORF 2280) (Fragment).
GN YCF2.
OS Oenothera villaricae.
OC Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC Myrtales; Onagraceae; Oenothera.
OX NCBI_TaxID=3941;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93169690; PubMed=8435856;
RA Nimzyk R., Shoendorf T., Hachtel W.;
RT "In-frame length mutations associated with short tandem repeats are
RT located in unassigned open reading frames of Oenothera chloroplast
RT DNA.";
RL Curr. Genet. 23:265-270(1993).
CC -----
CC -I- SIMILARITY: Belongs to the ycf2 family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X64615; CAA45896.1; -
DR PIR: S29796; S29796.
KM Chloroplast; Hypothetical protein.
FT NON TER
SQ SEQUENCE 630 AA; 72781 MW; 6AEPFF7DC75B0BA CRC64;

Query Match 4.3%; Score 172; DB 1; Length 630;
Best Local Similarity 20.9%; Pred. No. 0.087;
Matches 107; Conservative 82; Mismatches 168; Indels 154; Gaps 22;

QY 52 TSFTTFW-----MYALLGWTGYAVVMPVLV-----DYBEVLGKLGTYDADGCD 97
DB 100 SLYKWFYELGTSMKLTLLTYLLTCSAGSIAQDILSPGPDEQVLTSGLVENDSLV 159
QY 98 FDVDDAKVLGLKERSTSEPAVPEEAHPHTEPEEOVVEAEPQVIEDE--AKEQISL 154
DB 160 HGLSD--IVHGLLEBALVSSFTBEVEGTEBEVEGTEBEVEGTEBEVEGTEBEVAGT 217
QY 155 LHNEMVHA--EHVEG-----EDLQOEDGPTGSPQOEDDEFMATVDVDRFETLEPEV--SH 205
DB 218 EDEVEGTEBEVEGTEBEVEGTEBEVEGTEBEVEGTEBEVEGTEBEVEGTEBEVEGTE 276
QY 206 ETEHSYHVEYVSQDCNODMEEMSGENPDSSSEPVV--EDRRLHHTDDVY----- 256
DB 277 BEVGTEBEVEGTEBEVEGTEBEVEGTEBEVEGTEBEVEGTEKSSQFDNRVTLRLPK 336
QY 257 -----YQVE-----EQAYEP--LENEGIEIT----- 277
DB 337 PRNLDDQRLTYQKQKESHEEDDDDDDEVFAQKMLIEDLPSLWVSPRIIMHPDPLLD 396
QY 278 -EVAPEEDNPVEDSQV-----IVEVSLFVEE-----QOEVPEETNRK 316
DB 397 CEAEIPAEIPEEDDELPEDALTEVAVWGVEEGEADDEBDVLLAQOEDDELLEEDDE 456
QY 317 TDPEQKAKYKKKKPK---LNKPKDKTKKELDAEKL----- 351
DB 457 LDEEDELDEEPEEPEEDELHEEHEEHEEHEEHEEHEEHEEHEEHEEHEEHEEHEEHEE 516

```

```

QY 352 RKRGIT-----EEAVNAFELV-----RKPOSPRRARYGAACDD----- 388
Db 517 RKGDVFEVLSYPBATEISKEELLALNPKYRDAPKPKPRQKRWTKKQDKHYELLDRQ 576
QY 388 --LAEKR--RSNEVLRG--ALETYQEVASL 411
Db 577 RMLITKRLSKSNCGFFRSNTSESYQYLSNL 607

RESULT 15
L2M2_MOUSE STANDARD: PRT; 592 AA.
AC P21619;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DB Lamin B2.
GN LAMB2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RX NCBI_TaxID=10090;
RX [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91106216; PubMed=2102682;
RA Hoege T.H., Zatloukal K., Walzenegger I., Krohne G.;
RT "Characterization of a second highly conserved B-type lamin present
RT in cells previously thought to contain only a single B-type lamin.";
RL Chromosoma 99:379-390(1990).
RL [2]
RP ERRATUM.
RX MEDLINE=9139548; PubMed=2102440;
RA Hoege T.H., Zatloukal K., Walzenegger I., Krohne G.;
RL Chromosoma 100:67-69(1990).
CC -1- FUNCTION: Lamins are components of the nuclear lamina, a fibrous
CC layer on the nucleoplasmic side of the inner nuclear membrane,
CC which is thought to provide a framework for the nuclear envelope
CC and may also interact with chromatin.
CC -1- SUBCELLULAR LOCATION: Nucleoplasmic side of the inner nuclear
CC membrane.
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named Isoforms=2;
CC Name=B2;
CC IsoId=P21619-1; Sequence=Displayed;
CC Name=B3;
CC IsoId=P48680-1; Sequence=External;
CC Note=No experimental confirmation available;
CC -1- PM: B-TYPE LAMINS UNDERGO A SERIES OF MODIFICATIONS, SUCH AS
CC PHOSPHORYLATION AND PHOSPHORYLATION. INCREASED PHOSPHORYLATION OF
CC THE LAMINS OCCURS BEFORE ENVELOPE DISINTEGRATION AND PROBABLY
CC PLAYS A ROLE IN REGULATING LAMIN ASSOCIATIONS.
CC -1- MISCELLANEOUS: The structural integrity of the lamina is strictly
CC controlled by the cell cycle, as seen by the disintegration and
CC formation of the nuclear envelope in prophase and telophase,
CC respectively.
CC -1- SIMILARITY: Belongs to the intermediate filament family. THIS IS A
CC B TYPE LAMIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce)
CC or send an email to license@isb-sib.ch.
CC -----
CC EMBL; X54098; CAA38032.1; -.
CC PIR; B48345; B48315.
CC MGD; MGI:96796; Lmbb2.
CC GO; GO:0005638; C:lamin filament; IDA.
CC InterPro; IPR001664; IF_tail_C.
CC InterPro; IPR001322; IF_tail_C.
CC Pfam; PF00038; Filament; 1.

```

[illegible]

Search completed: May 19, 2004, 15:58:21
Job time : 22.2142 secs

17	691.5	17.2	215	11	Q920F7	Q920F7 mus musculus
18	550.5	13.7	133	11	Q9EB63	Q9EB63 mus musculus
19	535.5	13.3	147	11	Q9EB62	Q9EB62 mus musculus
20	516	12.8	245	11	Q8CH79	Q8CH79 mus musculus
21	435.5	10.8	210	4	Q9NR11	Q9NR11 homo sapien
22	420.5	10.5	210	6	Q28264	Q28264 canis famli
23	418	10.4	225	4	Q9NR10	Q9NR10 homo sapien
24	407.5	10.1	207	11	Q9EB64	Q9EB64 mus musculus
25	407.5	10.1	207	11	Q9CB06	Q9CB06 mus musculus
26	407.5	10.1	212	11	Q9D7J8	Q9d7J8 mus musculus
27	390.5	9.7	304	2	Q53792	Q53792 streptomyce
28	276	6.8	343	11	Q80VP9	Q80VP9 mus musculus
29	275	6.8	250	16	Q93H17	Q93H17 streptomyces
30	274.5	6.8	343	4	Q9UH33	Q9uh33 homo sapien
31	263.5	6.6	186	4	Q9NSU3	Q9nsu3 homo sapien
32	257	6.4	176	6	Q9N1E7	Q9n1E7 oryctolagus
33	251	6.2	176	11	Q9CUZ2	Q9cuZ2 mus musculus
34	250	6.2	49	11	Q9EC69	Q9ec69 mus musculus
35	217.5	5.4	230	4	Q8NAH3	Q8naH3 homo sapien
36	217.5	5.4	250	4	Q8N1I6	Q8niI6 homo sapien
37	217.5	5.4	280	4	Q8T6K3	Q8t6K3 homo sapien
38	217.5	5.4	347	4	Q96H00	Q96h00 homo sapien
39	216.5	5.4	110	13	Q91255	Q91255 petromyzon
40	207.5	5.2	1236	5	Q9GTX2	Q9gtx2 plasmodium
41	206	5.1	17352	2	Q95YX2	Q95yX2 procambilla
42	204.5	5.1	3111	5	Q9VH10	Q9vh10 drosophila
43	204	5.1	976	12	Q9PDN0	Q9pdn0 kaposi's sa
44	203	5.0	1233	5	Q8T56	Q8t56 plasmodium
45	199	4.5	1271	5	Q25860	Q25860 plasmodium

007 61 TAITGYWTSYAWWEDYNDVEEYIGRI

61 IALLGVWTSVAVVWFDLVDYEEVLGKLGIYDADGDGDFDVKVLLGLKERSTSEPAVP 120

01-OCT-2003 (TREMBlrel. 25, last annotation update)
 DB Aspartate-beta-hydroxylase.
 GN ASPH.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Ovary, and Uterus;
 RX MEDLINE=22354683; PubMed=12466851;
 RA The FANTOM Consortium,
 RT "The RIKEN Genome Exploration Research Group Phase I & II Team;
 RL "Analysis of the mouse transcriptome based on functional annotation of
 60,770 full-length cDNAs.";
 RT Nature 420:563-573 (2002).
 DR EMBL; AK030293; BAC26882.1; -
 DR MGD; MGI:1914166; Asph.
 DR InterPro; IPR007943; Asp-B-hydro.N.
 DR InterPro; IPR007803; Asp_Arg_Hydrox.
 DR InterPro; IPR008941; TPR-like.
 DR Pfam; PF05279; Asp-B-Hydro N: 1.
 DR Pfam; PF05118; Asp_Arg_Hydrox; 1.
 DR SMART; SM00028; TPR; 2.
 SQ SEQUENCE 741 AA; 83042 MW; 0660A6A5E34418C8 CRC64;

Query Match 79.8%; Score 3208.5; DB 11; Length 741;
 Best Local Similarity 80.6%; Pred. No. 3.2e-183;
 Matches 613; Conservative 41; Mismatches 71; Indels 37; Gaps 8;

1 MAQRKNAK-SSGNSSSSGSGSGS-----TSAGSSSPGARRETKHGHNKRGKGLSG 51
 1 MAPRKNKAGGSGGSSSSGSGSGSGSPSTGSSSSSPGARREAKGHNKRGKGLSG 60
 52 TSFPTFMVATALLGWTSTVAVWVMDLVYDEVYLGKGIYADGGDGDVDVDAKVLGLKE 111
 61 GSFTFMVATALLGWTSTVAVWVMDLVYDEVYLGKGIYADGGDGDVDVDAKVLGLKE 120
 112 RSTSEPAVPP-EEAPEPTPEEQVPEAEPONIDEAKEOIQSLIHMVAHEVGEBDLQ 170
 121 RSPSEKTFPPEEAEHFAELFEQAPREGADIONVDEYKEQIQSLIQSSTVHTD---DL- 175
 171 QEDGPTGEPQOEDEFLMATVDVDRFEETLEPEVSHEETSHYVEETVSODCNQDNEMM 230
 176 EADLAGEPQPEVEDEFLVTDSDRFEDELEPGVHEIEEDTYHVEDTASQNHNDMEEMT 235
 231 SEQNPSSSEPVVEDERLHDTDVTYQVEEQAVYEPLENEGIEITETVAAPEDNVED 290
 236 NEQNSDPSSEAVTDAGVLLPFAEEVRHQDYDE--PVYEPSEHEGVALS-----DNTIDD 287
 291 SQVIVEEVSIFPVEEQOEVPETNRKTDDEPQAKAVKKKKPKLNFKEDKTIKALDAAEK 350
 268 SSTISEINVAVEEQDTP-----VKKKKPKLNFKEDKTIKALDAAEK 333
 351 LRKRGKIEAVNAFKELVKKYPOSPPARYKGAQCEDDLAEKRSNEVLRAIETTYGEVAS 410
 334 LRKRGKIEAVNAFKELVKKYPOSPPARYKGAQCEDDLAEKRSNEVLRAIETTYGEVAS 393
 411 LPDVPALDKLSTLRBRDRQOFLGHRGSLTLTQRLVOLFNPNTSLKNDIGVYTLIGDN 470
 394 LPDVPALDKLSTLRBRDRQOFLGHRGSLTLTQRLVOLFNPNTSLKNDIGVYTLIGDN 453
 471 DNAKKVVEVLTVPNDGFAKVGFLKAKONKIAESIPLYKESIGSGDGTDDGRFYFH 530
 454 DSAKKVVEVLTVPNDGFAKVGFLKAKONKIAESIPLYKESIGSGDGTDDGRFYFH 513
 531 LGDMQVGNKEAYKMYELGHRGHPASVWORSILYNVGLKAOPWMTPKETGYTELKSTL 590
 514 LGDMQVGNKEAYKMYELGHRGHPASVWORSILYNVGLKAOPWMTPKETGYTELKSTL 573
 591 ERNWKILRDEGLAVMDAKGLFLPEDENLREKGMWSQFTLMQGRRENNCKGAPKTCCTL 650

DB 574 ERNWKILRDEGLAVMDAKGLFLPEDENLREKGMWSQFTLMQGRRENNCKGAPKTCCTL 633
 QY 651 LEKPEPTGCRGQIKYSIMHPGTHWPHGPTNCRIRLMHGLVYPEGGCKIRCANERT 710
 DB 634 LEKPEPTGCRGQIKYSIMHPGTHWPHGPTNCRIRLMHGLVYPEGGCKIRCANERT 693
 QY 711 WEEGKVLIFDPSPEHEVWODASSFRLIFIVDMHPELTPOGRSLPAI 758
 DB 694 WEEGKVLIFDPSPEHEVWODASSFRLIFIVDMHPELTPOGRSLPAI 741

RESULT 4

Q8BQKO PRELIMINARY; PRT; 725 AA.
 ID Q8BQKO
 AC Q8BQKO;
 DT 01-MAR-2003 (TREMBlrel. 23, Created)
 DT 01-MAR-2003 (TREMBlrel. 23, last sequence update)
 DT 01-OCT-2003 (TREMBlrel. 25, last annotation update)
 DE Aspartate-beta-hydroxylase.
 GN ASPH.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Body;
 RX MEDLINE=22354683; PubMed=12466851;
 RA The FANTOM Consortium,
 RT "The RIKEN Genome Exploration Research Group Phase I & II Team;
 RL "Analysis of the mouse transcriptome based on functional annotation of
 60,770 full-length cDNAs.";
 RT Nature 420:563-573 (2002).
 DR EMBL; AK049506; BAC33783.1; -
 DR MGD; MGI:1914166; Asph.
 DR InterPro; IPR007943; Asp-B-hydro.N.
 DR InterPro; IPR007803; Asp_Arg_Hydrox.
 DR InterPro; IPR008941; TPR-like.
 DR Pfam; PF05279; Asp-B-Hydro N: 1.
 DR Pfam; PF05118; Asp_Arg_Hydrox; 1.
 DR SMART; SM00028; TPR; 2.
 SQ SEQUENCE 725 AA; 81492 MW; F3CE979F8FC9C3D5 CRC64;

Query Match 79.0%; Score 3175.5; DB 11; Length 725;
 Best Local Similarity 79.8%; Pred. No. 2.9e-181;
 Matches 613; Conservative 41; Mismatches 61; Indels 53; Gaps 9;

1 MAQRKNAK-SSGNSSSSGSGSGS-----TSAGSSSPGARRETKHGHNKRGKGLSG 51
 1 MAPRKNKAGGSGGSSSSGSGSGSGSPSTGSSSSSPGARREAKGHNKRGKGLSG 60
 52 TSFPTFMVATALLGWTSTVAVWVMDLVYDEVYLGKGIYADGGDGDVDVDAKVLGLKE 111
 61 GSFTFMVATALLGWTSTVAVWVMDLVYDEVYLGKGIYADGGDGDVDVDAKVLGLKE 120
 112 RSTSEPAVPP-EEAPEPTPEEQVPEAEPONIDEAKEOIQSLIHMVAHEVGEBDLQ 170
 121 RSPSEKTFPPEEAEHFAELFEQAPREGADIONVDEYKEQIQSLIQSSTVHTD---DL- 175
 171 QEDGPTGEPQOEDEFLMATVDVDRFEETLEPEVSHEETSHYVEETVSODCNQDNEMM 230
 176 EADLAGEPQPEVEDEFLVTDSDRFEDELEPGVHEIEEDTYHVEDTASQNHNDMEEMT 235
 231 SEQNPSSSEPVVEDERLHDTDVTYQVEEQAVYEPLENEGIEITETVAAPEDNVED 290
 236 NEQNSDPSSEAVTDAGVLLPFAEEVRHQDYDE--PVYEPSEHEGVALS-----DNTIDD 271
 291 SQVIVEEVSIFPVEEQOEVPETNRKTDDEPQAKAVKKKKPKLNFKEDKTIKALDAAEK 350
 268 SSTISEINVAVEEQDTP-----VKKKKPKLNFKEDKTIKALDAAEK 317
 351 LRKRGKIEAVNAFKELVKKYPOSPPARYKGAQCEDDLAEKRSNEVLRAIETTYGEVAS 410

```

Db 318 LKRGKIEEAVAFELVAKYVQSPRRKRGKQCCEDDLAEKRSNEVLRRAIETQEAAD 377
Qy 411 LPDVPADLKLKSLKRSRQOFLGHRGSLTLQRLVOLFPNDTSLKNDLGVYLLIDN 470
Db 378 LPDAPDLVKSILKRSERQOFLGHRGSLTLQRLVOLFPNDTSLKNDLGVYLLIDN 437
Qy 471 DNAKKYEEVLSVTPNDGFAKVHGFLLKAONKIAESIPLYKEGIESDPTGDDGRFETH 530
Db 438 DSAKKYEEVLSVTPNDGFAKVHGFLLKAONKIAESIPLYKEGIESDPTGDDGRFETH 497
Qy 531 LGDAMQRYGNKEAYKWEELGHRGFASVWQSLYNNVGLKAQPMWTPKETYELVLSL 590
Db 498 LGDAMQRYGNKEAYKWEELGHRGFASVWQSLYNNVGLKAQPMWTPKETYELVLSL 557
Qy 591 ENNMKLIRDEGLAVMDKAKGLFLPEDENIREKDWMSQFTLMQOGRKNENACKGAPKTCIL 650
Db 558 ENNMKLIRDEGLAVMDKAKGLFLPEDENIREKDWMSQFTLMQOGRKNENACKGAPKTCAL 617
Qy 651 LEKPEETTCRRGQIKYSIMHPGTHVPHGTPTNCRMLHGLVTPKGGCKIRCANETRT 710
Db 618 LEKPEETTCRRGQIKYSIMHPGTHVPHGTPTNCRMLHGLVTPKGGCKIRCANETRT 677
Qy 711 WEEGKVLIFDSSFEEHFWQDASSFRLLIFIVDMHPELTPOORSLPAI 758
Db 678 WEEGKVLIFDSSFEEHFWQDASSFRLLIFIVDMHPELTPOORSLPAI 725

```

RESULT 5

Q8CBM2 PRELIMINARY; PRT; 658 AA.

```

ID Q8CBM2 AC Q8CBM2
DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Asparyl-beta-hydroxylase.
GN ASPH.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
CX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Urinary bladder;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA The RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs."
RL Nature 420:563-573(2002).
DR EMBL; AK035735; BAC29171.1;
DR MGD; MGI:191486; Asph.
DR InterPro; IPR007943; Asp-B-hydro_N.
DR InterPro; IPR007803; Asp_Arg_Hydrox.
DR InterPro; IPR008940; Prenyl_trans.
DR InterPro; IPR001440; TPR.
DR Pfam; PF05279; Asp-B-Hydro_N; 1.
DR Pfam; PF05118; Asp_Arg_Hydrox; 1.
DR SMART; SM00028; TPR; 2.
SQ SEQUENCE 658 AA; 75127 MW; 2BA0A5C8B06801C8 CRC64;

```

Query Match 73.3%; Score 2950; DB 11; Length 658;
 Best Local Similarity 80.7%; Pred. No. 7,3e-168;
 Matches 566; Conservative 38; Mismatches 53; Indels 44; Gaps 7;

```

Qy 59 MVALLLGVWTSVAVAVFDLVYEEVLTGLGYDADGDDPDVDKAVLLGLKRSSTSEPA 118
Db 1 MVALLLGVWTSVAVAVFDLVYEEVLTGLGYDADGDDPDVDKAVLLGLKRSSTSEPA 60
Qy 119 VPP-EEAAPHTEPEEQVPEAEPCNIDEAKEQIQSLHHEMVAHEHVGEDLQOEGPTG 177
Db 61 PPEEAEATTAIEBEQAPREGADIONVEDEVEEQIQSLQESVHTDH---DL-EEADLIG 115

```

```

Qy 178 EPOEDEFFLMTVDVDFETLPEPEVSHETESHVHEETVSOQCNOMEEMGEQENDP 237
Db 116 EPOEVEDFELVTIDSDRFEDLEFGVHEEIEDTYHVEDTASQNHPNMEEMTEQEN-- 173
Qy 238 SSEPVVEDERLHNDTQVYVEEQAVVELENEGIEITETVAPPENPVEDSQVYEE 297
Db 174 -----SEEVRRHODYDE-PYEESEHGVAI-S-----DNTIDDSIISEE 211
Qy 298 VSIFPVEEQEVPPEVETNRKTDPEQAKVKKKKKKPLINKFPDRTIKAEIDAAEKLRKSKI 357
Db 212 INVASVEEQDTP-----YKKKKRKLINKFPDRTIKAEIDAAEKLRKSKI 257
Qy 358 EEAVNAFEELVRYKTPQSPRARVYGAQCEDDIAEKRRSRNEVLRGAIEFYQVVASLPDVPAD 417
Db 258 EEAVNAFEELVRYKTPQSPRARVYGAQCEDDIAEKRRSRNEVLRGAIEFYQVAAADLPDAPTD 317
Qy 418 LKSLKRSRQOFLGHRGSLTLQRLVOLFPNDTSLKNDLGVYLLIDNNAKRYV 477
Db 318 LKSLKRSRQOFLGHRGSLTLQRLVOLFPNDTSLKNDLGVYLLIDNNAKRYV 377
Qy 478 EVELSVTPNDGFAKVHGFLLKAONKIAESIPLYKEGIESDPTGDDGRFETHLGDMQR 537
Db 378 EVELSVTPNDGFAKVHGFLLKAONKIAESIPLYKEGIESDPTGDDGRFETHLGDMQR 437
Qy 538 VGNKEAYKWEELGHRGFASVWQSLYNNVGLKAQPMWTPKETYELVKSLENNKLI 597
Db 438 VGNKEAYKWEELGHRGFASVWQSLYNNVGLKAQPMWTPKETYELVKSLENNKLI 497
Qy 598 RDEGLAVMDKAKGLFLPEDENIREKDWMSQFTLMQOGRKNENACKGAPKTCILLEKPEET 657
Db 498 RDEGLAVMDKAKGLFLPEDENIREKDWMSQFTLMQOGRKNENACKGAPKTCILEKSET 557
Qy 658 TGCRRGQIKYSIMHPGTHVPHGTPTNCRMLHGLVTPKGGCKIRCANETRTWEEGKVL 717
Db 558 TGCRRGQIKYSIMHPGTHVPHGTPTNCRMLHGLVTPKGGCKIRCANETRTWEEGKVL 617
Qy 718 IFDSSFEEHFWQDASSFRLLIFIVDMHPELTPOORSLPAI 758
Db 618 IFDSSFEEHFWQDASSFRLLIFIVDMHPELTPOORSLPAI 658

```

RESULT 6

Q9BQ66 PRELIMINARY; PRT; 669 AA.

```

ID Q9BQ66 AC Q9BQ66
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Aspartyl beta-hydroxylase.
GN ASPH.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
CX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=20564328; PubMed=10956665;
RX Dinchuk J.E., Henderson N.L., Burn T.C., Huber R., Ho S.P., Link J.,
RA O'Neil K.T., Focht R.J., Scully M.S., Hollis J.M., Hollis G.F.,
RA Friedman P.A.;
RT "Aspartyl beta-hydroxylase (Asph) and an evolutionarily conserved
RT isoform of asph missing the catalytic domain share exons with
RT junction."
RL J. Biol. Chem. 275:39543-39554(2000).
RN [2]

```

```

RP SEQUENCE FROM N.A.
RC Henderson N.L., Dinchuk J.E., Burn T.C., Hollis G.F., Friedman P.A.;
RA Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
RL EMBL; AF289215; AAG39913.1;
DR EMBL; AF289215; AAG39913.1; JOINED.
DR EMBL; AF289206; AAG39913.1; JOINED.
DR EMBL; AF289207; AAG39913.1; JOINED.
DR EMBL; AF289208; AAG39913.1; JOINED.

```


DR EMBL; AF289209; AAC39913.1; JOINED.
 DR EMBL; AF289210; AAC39913.1; JOINED.
 DR EMBL; AF289211; AAC39913.1; JOINED.
 DR EMBL; AF289212; AAC39913.1; JOINED.
 DR EMBL; AF289213; AAC39913.1; JOINED.
 DR EMBL; AF289214; AAC39913.1; JOINED.
 DR MGI;1914186; Asph.
 DR InterPro; IPR007943; Asp-B-Hydro_N.
 DR InterPro; IPR007803; Asp_Arg_Hydrox.
 DR InterPro; IPR001440; TPR.
 DR InterPro; IPR008941; TPR-like.
 DR Pfam; PF05279; Asp-B-Hydro_N; 1.
 DR Pfam; PF05118; Asp_Arg_Hydrox; 1.
 DR SEQUENCE 689 AA; 77319 MW; 9CB916DF109F432C CRC64;
 Query Match 72.6%; Score 2919.5; DB 11; Length 689;
 Best Local Similarity 74.7%; Pred. No. 5.2e-166;
 Matches 573; Conservative 39; Mismatches 68; Indels 87; Gaps 9;
 1 MAQRNANK-SSGNSSSSGSGSGS-----TSAGSSSPGARRETKHGKNGKRGKGLSG 51
 1 MAPRKNAKGSGGNSSSSGSGSGSGSGSPSTGSSSSSGSGPGRAR----- 43
 52 TSFPTFMVIALIGVWTSVAVMFVLDVYEBVIGKLGITYADGDGDDVDVDAKYLGLKE 111
 44 -----GKLGVDYADGDGDDVDVDAKYLGLKE 70
 112 RSTSEPAVPPEBAPHTPEBQVPEABPQNIEDAKEOIOSLHEMVAHEVGEEDLOQ 171
 71 RSPSERFPP-BAVTHALEBQAPEGADIQNVEDEVKEQIOSLQESVHIDH-----DL-E 124
 172 EDGTGTPQOBDDEFLMATVDVDRFETLEBPEVSHETESHVYEETVSQDCNQDMEMMS 231
 125 ADGAGBPOBQEVEDFLTVTDSDDFEDLEPQVHEETEDTVHVEDTASQNHPNMDEMTN 184
 232 EQENPDSEPEVBERLHHDITDVTYQVEBQAYEPELENEGIEITVTAPRENPVEDS 291
 185 EQESDSEAVTQDGVILPHAEVVRHQDYDE-PVYESEHEGVIS-----DNTTDS 236
 292 QVIVEEVSIFPVEEQOEVPEPTNKTDDPEOKAYKKKKKPKLANKPKTKTAKELDAEKL 351
 237 SIIEEELNVASVEEQOQTPP-----VKKKKPKLANKPKTKTAKELDAEKL 282
 352 RKRGKIEBAAVNAFELVRYKTPQSPRAYGKAQCEDDIAEKRSNEVLGAILETYQEVASL 411
 283 RKRGKIEBAAVNAFELVRYKTPQSPRAYGKAQCEDDIAEKRSNEVLRAILETYQEAADL 342
 412 PDVPAADLLKLSIKRSDROOFLGHRGSLTLTQRLVOLFENDTSLKNDLGVTLLGSDND 471
 343 PDAFTDLVKSLSIKRSDROOFLGHRGSLTLTQRLVOLFENDTSLKNDLGVTLLGSDND 402
 472 NAKKVEEVLSTPNDGFAKVHGYFLKAQNKIABSIPLYKEGIESGDPGTDDGRFYFHL 531
 403 SAKKVEEVLSTPNDGFAKVHGYFLKAQNKIABSIPLYKEGIESGDPGTDDGRFYFHL 462
 532 GDAMQRYGKEAAYVYELGKRGHFPASVWQSLYNNGLKAQPMWTPKEXGYELVKSLE 591
 463 GDAQRYGKEAAYVYELGKRGHFPASVWQSLYNNGLKAQPMWTPKEXGYELVKSLE 522
 592 RNWKLIDEGLAVNDKAKGLFLPEDENLREKQDMSQFTLMOQSRBNNAKGA PKTCTTL 651
 523 RNWKLIDEGLAVNDKAKGLFLPEDENLREKQDMSQFTLMOQSRBNNAKGA PKTCTTL 582
 652 EKPEPTGCRGQIKYSIMHGTWHPHTPTNCRIMHGLVYI PKEGCKIRCANETRTV 711
 583 EKSEETGCRGQIKYSIMHGTWHPHTPTNCRIMHGLVYI PKEGCKIRCANETRTV 642
 712 EBGVTLIFDPSFEHFWQDASSRLLITVVWHPBELTPQORRSIPAL 758
 643 EBGVTLIFDPSFEHFWQDASSRLLITVVWHPBELTPQORRSIPAL 689

Q9H2C4 ID Q9H2C4 PRELIMINARY; PRT; 313 AA.
 AC Q9H2C4;
 DT 01-MAR-2001 (Tremblrel. 16, Created)
 DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
 DT 01-JUN-2003 (Tremblrel. 24, Last annotation update)
 DE Aspartyl beta-hydroxylase 2.8 kb transcript.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Liver.
 RX MEDLINE=20564328; Pubmed=10956665;
 RA Dinchuk J.E., Henderson N.L., Burn T.C., Huber R., Ho S.P., Link J.,
 RA O'Neill K.T., Focht R.O., Scully M.S., Hollis J.M., Hollis G.F.,
 RA Friedman P.A.;
 RT "Aspartyl beta-hydroxylase (Asph) and an evolutionarily conserved
 RT isoform of asph missing the catalytic domain share exons with
 RT junction";
 RL J. Biol. Chem. 275:39543-39554(2000).
 DR EMBL; AF289489; AAC40811.1;
 DR InterPro; IPR007943; Asp-B-Hydro_N.
 DR Pfam; PF05279; Asp-B-Hydro_N; 1.
 DR SEQUENCE 313 AA; 34646 MW; 7885A18B81CD6D0D CRC64;
 Query Match 40.7%; Score 1636; DB 4; Length 313;
 Best Local Similarity 99.7%; Pred. No. 7.3e-90;
 Matches 312; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 1 MAQRNANKSSGNSSSSGSGSGSTAGSSSPGARRETKHGKNGKRGKGLSGTSEFTFMV 60
 1 MAQRNANKSSGNSSSSGSGSGSTAGSSSPGARRETKHGKNGKRGKGLSGTSEFTFMV 60
 61 IALLGWTSAVAVVFPVLDVYEEVLGKLGITYADGDGDDVDVDAKYLGLKERTSEBAPV 120
 61 IALLGWTSAVAVVFPVLDVYEEVLGKLGITYADGDGDDVDVDAKYLGLKERTSEBAPV 120
 121 PEBABPHTPEBQVPEABPQNIEDAKEOIOSLHEMVAHEVGEEDLOQEDGPTGEPQ 180
 121 PEBABPHTPEBQVPEABPQNIEDAKEOIOSLHEMVAHEVGEEDLOQEDGPTGEPQ 180
 181 QEDDEFMATVDVDRFETLEBPEVSHETESHVYEETVSQDCNQDMEMMSQENPDSS 240
 181 QEDDEFMATVDVDRFETLEBPEVSHETESHVYEETVSQDCNQDMEMMSQENPDSS 240
 241 PVVEDERLHHDITDVTYQVEBQAYEPELENEGIEITVTAPRENPVEDSQVIVEEVS 300
 241 PVVEDERLHHDITDVTYQVEBQAYEPELENEGIEITVTAPRENPVEDSQVIVEEVS 300
 301 FVVEEQOEVPEPT 313
 301 FVVEEQOEVPEPT 313
 RESULT 8
 Q9H291 ID Q9H291 PRELIMINARY; PRT; 299 AA.
 AC Q9H291;
 DT 01-MAR-2001 (Tremblrel. 16, Created)
 DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
 DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
 DE Junctate.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=20564329; Pubmed=11007777;
 RA Treves S., Feriotto G., Moccagatta L., Gambati R., Zorzato F.;
 RT "Molecular Cloning, Expression, Functional Characterization,

SEQUENCE 308 AA; 33142 MW; F208C2C31C595282 CRC64;

Query Match 24.8%; Score 998.5; DB 11; Length 308;
Best Local Similarity 64.3%; Pred. No. 8e-52;
Matches 207; Conservative 30; Mismatches 62; Indels 23; Gaps 7;

```

QY 1 MAQRKNAK-SGQNSSSSSGSGSGS-----TSAGSSSPGARRETKHGKNGKRGKGLSG 51
DB 1 MAQRKNAKGGGGSSSSSSSGSGSGSPSTGSSSGSSSPGARRETKHGKNGKRGKRGKGLSG 60
QY 52 TSFFTFMWIALIGVWTSVAVYVWFDLVYEEVLGKLGITYDADGGDFDVIDAKVLLGLKE 111
DB 61 GSFFTFMWIALIGVWTSVAVYVWFDLVYEEVLGKLGITYDADGGDFDVIDAKVLLGLKE 120
QY 112 RSTSPPAVPEEAEPHTEPEEQVPEAEPQNIIEDEAKQIQLLHMVYAAHVEGEDDQQ 171
DB 121 RSPSERTPP--EAETHALEEQAPGADIQNVEDEVKQIQLLQESVHTDH---DL-E 174
QY 172 EDGPTGEPQOEDDELMTDVDDRPTELEPEVSHETESHVHEVETVSQDCNQMEEMMS 231
DB 175 ADGLAGEPQPEVEDFLTVTSDRPEDLEPGTVHEIEDTYHVEDTASQNHENMEEMTN 234
QY 232 EQENPDSEPEVEDERLHHDITDVTYQVYEEQAVVEPLENGEITETVETAPPEDPVPS 291
DB 235 EQENPDSEAVTADAGVLLPHEAEVRAHQDYDE-PYIEPSEHGVGIS-----DNTIDS 286
QY 292 QVIEEVSIFPVEEQEVPPEPT 313
DB 287 SIISEINVASVEEQDTPPDT 308

```

RESULT 12

```

ID Q91WG6 PRELIMINARY; PRT; 292 AA.
AC Q91WG6;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
DE Similar to aspartate-beta-hydroxylase.
GN ASPH.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC Tissue=Kidney;
RA Strausberg R.;
RL Submitted (0CT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC015281; AAH15281.1; -.
DR MGD; MGI:1914186; Asph.
DR InterPro; IPR007943; Asp-B-Hydro_N.
DR Pfam; PF05279; Asp-B-Hydro_N; 1.
SQ SEQUENCE 292 AA; 31568 MW; B34470A46EC24B6D CRC64;

```

Query Match 24.2%; Score 972.5; DB 11; Length 292;
Best Local Similarity 62.7%; Pred. No. 2.7e-50;
Matches 202; Conservative 30; Mismatches 51; Indels 39; Gaps 8;

```

QY 1 MAQRKNAK-SGQNSSSSSGSGSGS-----TSAGSSSPGARRETKHGKNGKRGKGLSG 51
DB 1 MAQRKNAKGGGGSSSSSSSGSGSGSPSTGSSSGSSSPGARRETKHGKNGKRGKRGKGLSG 60
QY 52 TSFFTFMWIALIGVWTSVAVYVWFDLVYEEVLGKLGITYDADGGDFDVIDAKVLLGLKE 111
DB 61 GSFFTFMWIALIGVWTSVAVYVWFDLVYEEVLGKLGITYDADGGDFDVIDAKVLLGLKE 120
QY 112 RSTSPPAVPEEAEPHTEPEEQVPEAEPQNIIEDEAKQIQLLHMVYAAHVEGEDDQQ 171
DB 121 RSPSERTPP--EAETHALEEQAPGADIQNVEDEVKQIQLLQESVHTDH---DL-E 174
QY 172 EDGPTGEPQOEDDELMTDVDDRPTELEPEVSHETESHVHEVETVSQDCNQMEEMMS 231

```

```

DB 175 ADGLAGEPQPEVEDFLTVTSDRFEDELPGTVHEIEDTYHVEDTASQNHENMEEMTN 234
QY 232 EQENPDSEPEVEDERLHHDITDVTYQVYEEQAVVEPLENGEITETVETAPPEDPVPS 291
DB 235 EQEN-----SEVYRHQDYDE-PYIEPSEHGVGIS-----DNTIDS 270
QY 292 QVIEEVSIFPVEEQEVPPEPT 313
DB 271 SIISEINVASVEEQDTPPDT 292

```

RESULT 13

```

ID Q920F9 PRELIMINARY; PRT; 270 AA.
AC Q920F9;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
DE Cardiac junction 1.
GN ASPH.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALE/c;
RA Hong C.-S., Kim D.H.;
RT "Cloning of mouse junction homologs.";
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF302653; AAL09319.2; -.
DR PIR; JC7792; JC7792.
DR MGD; MGI:1914186; Asph.
DR InterPro; IPR007943; Asp-B-Hydro_N.
DR Pfam; PF05279; Asp-B-Hydro_N; 1.
SQ SEQUENCE 270 AA; 29949 MW; 70BB813DC01B51701 CRC64;

```

Query Match 22.1%; Score 890; DB 11; Length 270;
Best Local Similarity 63.9%; Pred. No. 2e-45;
Matches 179; Conservative 30; Mismatches 57; Indels 14; Gaps 5;

```

QY 34 RETHGKNGKRGKGSSTGTFMFVIALIGVWTSVAVYVWFDLVYEEVLGKLGITYDAD 93
DB 5 KEAKHGKNGKRGKGSSTGTFMFVIALIGVWTSVAVYVWFDLVYEEVLGKLGITYDAD 64
QY 94 GDGDFVDYDAKVLGLKERSSTSEPAVPEEAEPHTEPEEQVPEAEPQNIIEDEAKQIQL 153
DB 65 GDGDFVDYDAKVLGLKERSPSERTPP--EAETHALEEQAPGADIQNVEDEVKQIQL 123
QY 154 LHMVYAAHVEGEDDQQEDGPTGEPQOEDDELMTDVDDRPTELEPEVSHETESHV 213
DB 124 LQESVHTDH---DL-EADGLAGEPQPEVEDFLTVTSDRPEDLEPGTVHEIEDTYH 178
QY 214 VEETVSQDCNQMEEMMSQENPDSEPEVEDERLHHDITDVTYQVYEEQAVVEPLENGE 273
DB 179 VEDTASQNHENMEEMNSQENPDSEPAVYDAGVLLPHEAEVRAHQDYDE-PYIEPSEH 237
QY 274 IEITEVTAPEEDNPVEDSQVIEEVSIFPVEEQEVPPEPT 313
DB 238 VEIS-----DNTIDSSIISEINVASVEEQDTPPDT 270

```

RESULT 14

```

ID Q93178 PRELIMINARY; PRT; 872 AA.
AC Q93178; Q93876;
DT 01-FEB-1997 (TREMBLrel. 02, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE K09A9.6 protein.
GN Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;

```

OC Rhabditiidae; Pelodierinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Swinburne J.;
 RL Submitted (Aug-1996) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94150718; PubMed=7906398;
 RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
 RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
 RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
 RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
 RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
 RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
 RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,
 RA Smalton N., Smith A., Sonhammer E., Straden R., Sulston J.,
 RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
 RA Watson A., Weinstock L., Wilkinson-Spirot J., Woldman P.,
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 RT elegans";
 RL Nature 368:32-38 (1994).
 DR EMBL; Z79601; CAB01887.1; -
 DR EMBL; Z79596; CAB01887.1; JOINED.
 DR EMBL; Z79596; CAB01859.1; -
 DR EMBL; Z79601; CAB01859.1; JOINED.
 DR PIR; T18861; T18861.
 DR WormPep; K09A9.6; CE11982.
 DR InterPro; IPR007803; Asp_Arg_Hydrol.
 DR InterPro; IPR008940; PreNyl_Ctrans.
 DR InterPro; IPR001440; TPR.
 DR Pfam; PF05118; Asp_Arg_Hydrol; 1.
 SQ SEQUENCE 872 AA; 99456 MW; 8CA3E03489A9848B CRC64;

Query Match 19.8%; Score 798; DB 5; Length 872;
 Best Local Similarity 26.8%; Pred. No. 3.1e-39;

Matches 191; Conservative 144; Mismatches 278; Indels 100; Gaps 15;

89 IYDADGDDPVDVDAKYLGLKERSTSEPAVPEEAEFHTPEPEQ-----VPVEAPQ 141
 Db VEDDDDDDDDDDDVE-----APAPQEPKQKQAHNQREKKQKNNKQPVKXEPD 255
 Qy 142 NIDPEAKEQIQLHEWVAHVEGEDLQOEDEG-----PTGEQOQDEDFMATVDYDR 135
 Db 256 LDDDDDEDDDDDDGDDDDDEKDDGNAEKDDGDDDDDDDDGDDDEDEKNNKTSYEAK 315
 Qy 196 FETL-----EPEVSHETE--HSYVET-----VSQDC-NQDME- 228
 Db 316 SDKVEKKNQDDEPFVSHREAOQLRQHRLRESKRDNRPRQGNRECIHDDCNRRESLKP 375
 Qy 229 ---MMSEQENPDSSEPFVDEDERLHNDTDDVYQVVEQAQVYEBLENGIEITEVTAPE 284
 Db 376 RKSLLVTKKTKSVERILDDDEFDDDDDD-----EDDSEVANK-----N 415
 Qy 285 DNPVEDSQVIVVEKSTIFVVEEQEVPEPTNRKTDDEPQAKVKKKKKXKLINKEDKTKAE 344
 Db 416 DREDDDDDDVDERIS-----DRDSSSSYKRAIITTKKEIGFRDI 454
 Qy 345 LDAEKRKRKGKLEAVNAFAELVRYKQSPRARVYGAQCEDDLAEGKRSNEVLGRIET 404
 Db 455 LDRADNIVEKHQYEAHELDHVAIVYPASTRAYFGAKRAYDTRGELEADETDRDAIET 514
 Qy 405 YQEVASLDDVPADLLKLSLRSDRQQLGHRGSLITLQLVQLFPNDTSLKNDLGAVY 464
 Db 515 YEKILQSGVVDALFRQAQRLIKTRFRGLKHTLPAHYFYDFRFEELNLQTFDAISF 574
 Qy 465 LILGDNNAKKYVEVSVTPNDGPAKHYGFILKA-QNKIAEPIFLKGISSGDPGT 523
 Db 575 VMMKRYEDARTVKNVLANDEPNHVALAYGYLLKADHDVVEQVALMRKSLKNADEIT 634
 Qy 524 DGRFFPHGDAMQGVGNK-EAYKWEELGHRGHPASVWQSLYVNGIAKQPMWTPKETG 582
 Db 635 DKRYTGLGGLTTLGKKSBDADVYQKAAQMGVMTAQGRSLYINIBGLTGRAMWAMQTP 694

Qy 583 YTELVSLERMNKLIDEGIAVMDKAGLFLPDENILREKQWSQFTLMOQGRNENACK 642
 Db 695 YSKFLKTVERQWATIRIQEGWEVLKDCSDCLDHNQOLVIDGQKFFPIMEQNFIKSSCE 754
 Qy 643 GAKTCTLLEKFPETTCGRCGOIKYSIMPHGTHVHTGPTNRLKMHGLVYPKXECKI 702
 Db 755 RMQOTLLIQEFPAASNASKSDMHLSSGASILPHCGPTNHLQMHGLVSPSE-ARI 813
 Qy 703 RCANERTVEGKVLIFDDSFHEHVMOD---ASSFRLIFVDVWHPELTPQOR 752
 Db 814 RVNERTKMRSGKFIYDVSFEHLEQFDGASSSFRVLVTQLQMHPEVQHNOR 866

RESULT 15

Q920F8 PRELIMINARY; PRT; 259 AA.
 AC Q920F8;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, last annotation update)
 DE Cardiac junctate 2.
 GN ASPH OR 3110001L23PRK.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BALB/C;
 RA Hong C.-S., Kim D.H.;
 RT "Cloning of mouse junction homologs."
 RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF302654; AAL09320.1; -
 DR MGD; MGI:1914186; Asph.
 DR InterPro; IPR007943; Asp-B-Hydro_N.
 DR Pfam; PF05279; Asp-B-Hydro_N; 1.
 SQ SEQUENCE 259 AA; 28455 MW; A6740CC6CF199E093 CRC64;

Query Match 19.4%; Score 779.5; DB 11; Length 259;

Best Local Similarity 55.4%; Pred. No. 7.5e-39;

Matches 163; Conservative 27; Mismatches 51; Indels 53; Gaps 6;

Qy 34 RETHGHHKNGKRGKGLSGTSFFFTWENVIALGVWTSYAAVWFDLVYEEVLGKIGTDAD 93
 Db 5 KEAKHGHHKNGKRGKGLSGTSFFFTWENVIALGVWTSYAAVWFDLVYEEVLGKIGTDAD 64
 Qy 94 GDDGFDVDDAKVYL-----GLKERSTSEPAVPEEAEFHTPEPEQVPEAE 139
 Db 65 GDDGFDVDDAKVYLBEGPGGLAKKRYAKGLKERSPSERTPP-BATHALEBOAP----- 119
 Qy 140 PNIIDPEAKEQIQLHEWVAHVEGEDLQOEDEGPTGEBOQDEDFMATVDYDRPETL 199
 Db 120 -----EGADL-BADGLAGEQPRVEDFLVTDSDDREPDL 153
 Qy 200 EPEVSHETESHVVEETYSQDCNQDMEEMSEQENPDSEPFVDEDERLHNDTDDVYQV 259
 Db 154 EPGTVHELEDTYHVEPTASQNHNDMEETNQSNDPBEAVTDAVLLPFAEEVNHOD 213
 Qy 260 YEQAVYEPLENGIEITEVTAPEPDNPVDSQVIVVEVSIIPVVEEQEVPEPT 313
 Db 214 YDE-PYVESEHGEVLEIS-----DVTIDSSITSEINAVASVEBQDTPPT 259

Search completed: May 19, 2004, 15:59:50
 Job time : 76.0139 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 19, 2004, 15:50:28 ; Search time 4.0076 Seconds
(without alignments)
2185.588 Million cell updates/sec

Title: US-09-903-199-4

Perfect score: 139

Sequence: 1 CDXXXKXKXGNGXCDXXCNNAACXXDXDC 31

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

A_Geneseq_23Jan04:*
1: geneseqp19808:*
2: geneseqp19908:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	115	82.7	31	4	AAB83920 Amino aci
2	115	82.7	31	6	ABg72366 Epidermal
3	115	82.7	31	6	ADA00641 EGF-Like
4	105	75.5	2703	4	ABB60266 Drosophil
5	85	61.2	2531	7	ADBE3713 Rat Prote
6	85	61.2	2531	7	ADBE3705 Rat Prote
7	85	61.2	2531	7	ADBE3709 Rat Prote
8	85	61.2	2531	7	ADBE3701 Rat Prote
9	84	60.4	2444	5	ABB07821 Constitut
10	84	60.4	2469	5	AAE18207 Human MOL
11	84	60.4	2469	7	ADD18192 Human Not
12	84	60.4	2471	2	AAO27065 Human Not
13	84	60.4	2471	2	AAV06816 Human Not
14	84	60.4	2471	6	AAg79774 Human Not
15	84	60.4	2471	6	ABP72572 Human Not
16	84	60.4	2471	6	ABR61831 Human Not
17	84	60.4	2471	7	ABR61760 Human Not
18	84	60.4	2471	7	ADBE38245 Human Pro
19	84	60.4	2471	7	ADBE3707 Human Pro
20	84	60.4	2471	7	ADBE3703 Human Pro
21	84	60.4	2471	7	ADBE3715 Human Pro
22	84	60.4	2471	7	ADBE38243 Rat Prote
23	84	60.4	2471	7	ADBE3711 Human Pro
24	84	60.4	2556	2	AAO27066 Human Not
25	84	60.4	2556	6	ABg70518 Human pol

26	84	60.4	2556	6	AAg79773 Human Not
27	84	60.4	2556	6	ABP72571 Human Not
28	84	60.4	2556	6	ABR61830 Human Not
29	84	60.4	2556	6	ABR61759 Human Not
30	83	59.7	1872	2	AAW68510 Partial h
31	83	59.7	2321	2	AAW49698 Human Not
32	79	56.8	1764	2	AAW95557 Mus muscu
33	78	56.1	1846	2	ABU11844 Human MDD
34	78	56.1	946	6	ADA54824 Human pro
35	78	56.1	928	7	ABR61376 Human Glc
36	78	56.1	928	7	ADD27812 Human Glc
37	78	56.1	928	7	ABW01488 Human Glc
38	78	56.1	928	7	ABW01537 Human Glc
39	78	56.1	1199	7	ADD27810 Soluble h
40	78	56.1	1199	7	ABW01487 N-acetyl1g
41	78	56.1	1199	7	ABW01536 N-acetyl1g
42	78	56.1	1256	5	AAE25290 Human nuc
43	78	56.1	1459	6	ABU07381 Human pro
44	77	55.4	1078	2	AAr28963 Notch hn3
45	77	55.4	1078	2	AAO27058 Human Not

ALIGNMENTS

RESULT 1
AAB83920
ID AAB83920 standard; peptide; 31 AA.

AC AAB83920;

DT 23-JUL-2001 (first entry)

DE Amino acid sequence of consensus epidermal growth factor-like domain.

XX Epidermal growth factor-like domain; EGF-like domain; cancer;

KW human aspartyl beta-hydroxylase; HAAH; malignant neoplasm; tumour.

XX Synthetic.

OS Key Location/Qualifiers

FT Misc-difference 1..31 /note="Xaa represents an unspecified residue"

XX WO200135102-A2.

XX 17-MAY-2001.

XX 08-NOV-2000; 2000WO-US030738.

XX 08-NOV-1999; 99US-00436184.

XX (RHOD-) RHODE ISLAND HOSPITAL LIFESPAN PARTNER.

XX Wands JR, De La Monte SM, Ince N, Carlson RI;

XX WPI; 2001-329171/34.

XX Diagnosing malignant neoplasm in a mammal, involves contacting mammalian

XX sample with antibody that binds to human aspartyl beta-hydroxylase

XX polypeptide to form antigen-antibody complex and detecting the complex.

XX Disclosure; Page 7; 76pp; English.

XX The present sequence represents a consensus epidermal growth factor (EGF)

XX -like cysteine-rich repeat. EGF-like domains of polypeptides are

XX hydroxylated by a human aspartyl (asparaginyl) beta-hydroxylase (HAAH)

XX enzyme. HAAH is used in the method of the invention. The specification

XX describes a method for diagnosing a malignant neoplasm in a mammal. The

XX method comprises contacting a body fluid with an antibody which binds to

XX HAAH polypeptide under complex forming conditions, and detecting the

XX antigen-antibody complex. The method is useful for diagnosing and

XX prognosing a malignant neoplasm in a bodily fluid e.g. central nervous

```
CC system(CNS)-derived body fluid, blood, serum, urine, saliva, sputum,  
CC lung effusion), and ascites fluid of mammal, where the neoplasm is derived  
CC from endodermal tissue and is selected from colon cancer, breast cancer,  
CC pancreatic cancer, liver cancer, cancer of bile ducts, and cancer of the  
CC CNS. HAHM antibodies, linked to a cytotoxic agent, are useful for killing  
CX tumour cells
```

```
SQ      Sequence 31 AA;  
  
Query Match                               82.7%; Score 115; DB 4; Length 31;  
Best Local Similarity    100.0%; Pred.No. 7.le-08;  
Matches   31; Conservative     0; Mismatches       0; Indels        0; Gaps          0;
```

```
OY      1 CDXXXCXXKXGNGCDXXCNNAACXXDXDC 31  
         |||||  
Db       1 CDXXXCXXKXGNGCDXXCNNAACXXDXDC 31
```

```
RESULT 2  
ABG72366  
ID      ABG72366 standard; peptide; 31 AA.  
XX  
AC      ABG72366;  
XX  
DT      06-FEB-2003 (first entry)  
XX  
DE      Epidermal growth factor-like cyteine-rich repeat consensus sequence.
```

```
KW Aspartyl[asparaginy] beta-hydroxylase; HAHM; cytosstatic;  
KW immunoglobulin; antibody; neoplasia; tumor; FB50; 86A; 5C7; 19B;  
KW brain tumor; glioma; glioblastoma; astrocytoma; haemangioma;  
KW pancreatic carcinoma; colon cancer; breast cancer; pancreatic cancer;  
KW liver cancer; cancer of the bile ducts; primary malignant CNS neoplasia;  
KW metastatic CNS neoplasia; BGF; NOTCH;  
KM Epidermal growth factor-like cyteine-rich repeat consensus sequence.
```

```
OS Synthetic.  
XX  
FH Key Location/Qualifiers  
FT Misc-difference 3..15 /label= UNKNOWN  
FT FT /label= UNKNOWN  
FT Misc-difference 7..18 /label= UNKNOWN  
FT FT /label= UNKNOWN  
FT Misc-difference 10 /label= UNKNOWN  
FT FT /label= UNKNOWN  
FT Misc-difference 14 /label= UNKNOWN  
FT FT /label= UNKNOWN  
FT Misc-difference 17..18 /label= UNKNOWN  
FT FT /label= UNKNOWN  
FT Misc-difference 25..26 /label= UNKNOWN  
FT FT /label= UNKNOWN  
FT Misc-difference 29 /label= UNKNOWN  
FX  
FN US200210559-A1.  
XX  
PD 15-AUG-2002.  
XX  
PF 17-MAY-2001; 2001US-00859604.  
XX  
PR 08-NOV-1999; 99US-00436184.  
XX  
PA ('WAND') WANDS J R.  
PA (DMON/) DE LA MONTE S M.  
PA (DEUT/) DEUTCH A H.  
PA (GHAN/) GHANBARI H A.
```

```
P1 Wands JR, De La Monte SM, Deutsch AH, Ghanbari HA;  
DR WP1: 2003-066676/06.  
XX  
XT Diagnosing neoplasm and inhibiting tumor growth, by contacting tissue of  
XT mammal with detectably-labeled antibody which binds to human aspartyl
```

PT	(asparaginyl) beta-hydroxylase.
XX	
PS	Disclosure; page 5; 34pp; English.
XX	
CC	The invention relates to diagnosing a neoplasm and inhibiting tumour growth in a mammal, using an antibody that binds to human aspartyl (asparaginyl) beta-hydroxylase (HAH). Diagnosing a neoplasm comprises contacting a tissue with a detectably-labelled antibody where an increase in level of antibody binding at tissue site compared to the level of binding to normal non-neoplastic tissue indicates the presence of a neoplasm at the tissue site. Inhibiting tumour growth in mammal involves administering the antibody conjugated to a cytotoxic agent to a mammal. Also included are a method of conferring an immune response to a tumour cell in a mammal, by administering the antibody, a method of inducing an HAH-specific immune response in a mammal, by administering to the mammal an HAH polypeptide (or a polynucleotide composition encoding the polypeptide, or its degenerate variant), a fragment of HAH comprising an extracellular domain and lacking a cytoplasmic domain of HAH, an antibody or its fragment which binds to HAH (where the antibody is PB50, 86A, 507 or 19B), a hybridoma cell line chosen from hybridoma PB50, HA386A, HA15CA and HA219B, and a fragment of HAH which lacks enzymatic activity or alpha-ketoglutarate binding domain and epidermal growth factor (EGF)-like domain. The methods are useful for diagnosing neoplasm in a mammal, inhibiting tumour growth in a mammal, conferring an immune response to a brain tumour cell (e.g. glioma, glioblastoma, astrocytoma or haemangioma) in a mammal, for conferring immune response to a pancreatic carcinoma cell and for inducing a HAH-specific immune response in a mammal. The method is useful for diagnosing malignant neoplasms derived from endodermal tissue, e.g. colon cancer, breast cancer, pancreatic cancer, liver cancer and cancer of the bile ducts. The method is also useful for diagnosing neoplasms of central nervous system (CNS) e.g. primary malignant CNS neoplasms and for diagnosing brain tumours e.g. glioma, glioblastoma, astrocytoma or haemangioma. The present sequence represents an Epidermal growth factor (EGF)-like cysteine-rich repeat consensus sequence (e.g. from the NOTCH protein) which is hydroxylated by human HAH
XX	
XX	
SQ	Sequence 31 AA;
CC	
CC	Query Match 92.7%; Score 115; DB 6; Length 31;
CC	Best Local Similarity 100.0%; Pred.No. 7.1e-08;
CC	Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY	1 CDXXXXXXXGXGKCDXXCNNAACXXGXGXC 31 1 CDXXXXXXXGXGKCDXXCNNAACXXGXGXC 31
DB	
RESULT 3	
ID	ADA00641 standard; peptide; 31 AA.
XX	
NC	ADA00641;
XX	
DT	06-NOV-2003 (first entry)
XX	
DE	EGF-like cysteine-rich repeat consensus sequence.
XX	
XX	Tumour growth inhibition; human aspartyl (asparaginyl) beta-hydroxylase; HAH hydroxylation; NOTCH polypeptide; EGF-like repeat; tumour cell; malignant growth factor-like repeat; EGF-like repeat; tumour cell; malignant neoplasm; colon cancer; breast cancer; pancreatic cancer; liver cancer; cancer of the bile duct; cancer the central nervous system; CNS; cytostatic; EGF-like cysteine-rich repeat.
XX	
XX	Synthetic.
XX	
FT	Key Location/Qualifiers
FT	Misc-difference 3..5
FT	/label= Unknown
FT	Misc-difference 7..8
FT	/label= Unknown

FT Misc-difference 10
 FT /label= Unknown
 FT Misc-difference 14
 FT /label= Unknown
 FT Misc-difference 17.18
 FT /label= Unknown
 FT Misc-difference 25.26
 FT /label= Unknown
 FT Misc-difference 29
 FT /label= Unknown
 FT US2003031670-A1.
 PN 13-FEB-2003.
 PD 08-NOV-1999; 99US-00436184.
 XX 08-NOV-1999; 99US-00436184.
 PR 08-NOV-1999; 99US-00436184.
 XX
 XX (WAND/) WANDS J R.
 PA (DMON/) DE LA MONTE S M.
 PA (INCE/) INCE N.
 PA (CARL/) CARLSON R I.
 PI Wands JR, De La Monte SM, Ince N, Carlson RI;
 PI WPI; 2003-605701/57.
 DR
 XX
 PT Inhibiting tumor growth or killing tumor cells (e.g. cancer of the colon,
 PT breast, pancreatic, liver or the central nervous system), by
 PT administering an inhibitor of the human aspartyl (asparaginyl) beta-
 PT hydroxylase.
 XX
 PS Disclosure; Page 4; 30pp; English.
 XX
 CC The present invention relates to a method for inhibiting tumour growth in
 CC a mammal. The method comprises administering to the mammal a compound,
 CC which inhibits the expression or enzymatic activity of a human aspartyl
 CC (asparaginyl) beta-hydroxylase (HAH). The compound may inhibit HAH
 CC hydroxylation of a NOTCH polypeptide. In particular, the compound may
 CC inhibit hydroxylation of an epidermal growth factor (EGF)-like repeat
 CC sequence in a NOTCH polypeptide. The methods are useful for inhibiting
 CC tumour growth or killing tumour cells, or for diagnosing or
 CC prognosticating a malignant neoplasm. In particular, the tumour or
 CC neoplasm is colon cancer, breast cancer, pancreatic cancer, liver cancer,
 CC cancer of the bile ducts, or cancer or tumour of the central nervous
 CC system (CNS). The present sequence represents an EGF-like cysteine-rich
 CC repeat consensus sequence.
 CC
 SQ Sequence 31 AA;
 Query Match 82.7%; Score 115; DB 6; Length 31;
 Best Local Similarity 100.0%; Pred.No. 7.1e-08;
 Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 CDXXXXCXXKXGKXCDXXCNNAACXXDXGDC 31
 Db 1 CDXXXXCXXKXGKXCDXXCNNAACXXDXGDC 31
 RESULT 4
 ABB60266
 ID ABB60266 standard; protein; 2703 AA.
 XX
 AC ABB60266;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 XX Drosophila melanogaster polypeptide SEQ ID NO 7590.
 DE
 XX
 XX Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical.
 KW
 PA

OS Drosophila melanogaster.
 PN WO200171042-A2.
 XX
 XX 27-SEP-2001.
 PD
 XX
 XX 23-MAR-2001; 2001WO-US009231.
 PF
 XX
 PR 23-MAR-2000; 2000US-0191637P.
 PR 11-JUL-2000; 2000US-00614150.
 XX
 XX (PEKE) PE CORP NY.
 PA
 PA Ventier JC, Adams M, Li PWD, Myers EM;
 PI WPI; 2001-656860/75.
 DR N-PSDB; ABL04369.
 DR
 XX
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions.
 XX
 PS Disclosure; SEQ ID NO 7590; 21pp + Sequence Listing; English.
 XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins (ABBS7737-
 CC ABB72072). The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
 CC
 SQ Sequence 2703 AA;
 Query Match 75.5%; Score 105; DB 4; Length 2703;
 Best Local Similarity 54.8%; Pred.No. 5.3e-05;
 Matches 17; Conservative 0; Mismatches 14; Indels 0; Gaps 0;
 QY 1 CDXXXXCXXKXGKXCDXXCNNAACXXDXGDC 31
 Db 1482 CDKRGCTERQNGICDSDNCTYACNFDGND 1512
 RESULT 5
 ADB63713
 ID ADB63713 standard; protein; 2531 AA.
 XX
 AC ADB63713;
 XX
 DT 29-JAN-2004 (first entry)
 DT
 DE Rat Protein CAA40667, SEQ ID NO 9657.
 XX
 XX Rat; pain; neuronal tissue; gene therapy; spinal segmental nerve injury;
 KW chronic constriction injury; CCI; spared nerve injury; SNI; Chung.
 KW
 OS Rattus norvegicus.
 XX
 XX WO2003016475-A2.
 PN
 PD 27-FEB-2003.
 PD
 XX
 XX 14-AUG-2002; 2002WO-US025765.
 PF
 XX
 PR 14-AUG-2001; 2001US-0312147P.
 PR 01-NOV-2001; 2001US-0346382P.
 PR 26-NOV-2001; 2001US-033347P.
 XX
 XX (GEHO) GEN HOSPITAL CORP.
 PA (FARB) BAYER AG.

XX Woolf C, D'urso D, Befort K, Costigan M;
 PI WPI: 2003-268312/26.
 DR GENBANK, CAA40667.
 XX
 XX New composition comprising two or more isolated polypeptides, useful for
 PT preparing a medicament for treating pain in an animal.
 XX
 PS Claim 1, Page: 1017pp; English.

CC The invention discloses a composition comprising two or more isolated rat
 CC or human polynucleotides or a polynucleotide which represents a fragment,
 CC derivative or allelic variation of the nucleic acid sequence. Also
 CC claimed are a vector comprising the novel polynucleotide, a host cell
 CC comprising the vector, a method for identifying a nucleotide sequence
 CC which is differentially regulated in an animal subjected to pain and a
 CC kit to perform the method, an array, a method for identifying an agent
 CC that increases or decreases the expression of the polynucleotide sequence
 CC that is differentially expressed in neuronal tissue of a first animal
 CC subjected to pain, a method for identifying a compound which regulates
 CC the expression of a polynucleotide sequence which is differentially
 CC expressed in an animal subjected to pain, a method for identifying a
 CC compound that regulates the activity of one or more of the
 CC polynucleotides, a method for producing a pharmaceutical composition, a
 CC method for identifying a compound or small molecule that regulates the
 CC activity in an animal of one or more of the polypeptides given in the
 CC specification, a method for identifying a compound useful in treating
 CC pain and a pharmaceutical composition comprising the one or more
 CC polypeptides or their antibodies. The polynucleotide or the compound that
 CC modulates its activity is useful for preparing a medicament for treating
 CC pain (e.g. spinal segmental nerve injury (Chung), chronic constriction
 CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
 CC therapy). The sequence presented is a rat protein (shown in Table 2 of
 CC the specification) which is differentially expressed during pain. Note:
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic form directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.

CC Sequence 2531 AA;
 XX
 XX

Query Match 61.2%; Score 85; DB 7; Length 2531;
 Best Local Similarity 45.2%; Pred. No. 0.017;
 Matches 14; Conservative 2; Mismatches 15; Indels 0; Gaps 0;

QY 1 CDXXXCXXKXGNGCXDXXCNNNAACXXDXDC 31
 Db 1449 CELPECQDAGNKKVCNLCQNNHACGWDGDC 1479

RESULT 6
 ADE63705
 ID ADE63705 standard; protein; 2531 AA.
 XX
 AC ADE63705;
 XX
 DT 29-JAN-2004 (first entry)
 XX
 DE Rat Protein CAA40667, SEQ ID NO 9649.
 XX
 KW Rat, pain; neuronal tissue; gene therapy; spinal segmental nerve injury;
 KW chronic constriction injury; CCI; spared nerve injury; SNI; Chung.
 XX
 OS Rattus norvegicus.
 XX
 PN MO2003016475-A2.
 XX
 PD 27-FEB-2003.
 XX
 PF 14-AUG-2002; 2002WO-US025765.
 XX
 PR 14-AUG-2001; 2001US-0312147P.
 PR 01-NOV-2001; 2001US-0346382P.

ER 26-NOV-2001; 2001US-0333347P.
 XX
 XX (GENO) GEN HOSPITAL CORP.
 PA (FARB) BAYER AG.
 XX
 XX
 PI Woolf C, D'urso D, Befort K, Costigan M;
 XX WPI: 2003-268312/26.
 DR GENBANK, CAA40667.
 XX
 XX New composition comprising two or more isolated polypeptides, useful for
 PT preparing a medicament for treating pain in an animal.
 XX
 PS Claim 1, Page: 1017pp; English.

CC The invention discloses a composition comprising two or more isolated rat
 CC or human polynucleotides or a polynucleotide which represents a fragment,
 CC derivative or allelic variation of the nucleic acid sequence. Also
 CC claimed are a vector comprising the novel polynucleotide, a host cell
 CC comprising the vector, a method for identifying a nucleotide sequence
 CC which is differentially regulated in an animal subjected to pain and a
 CC kit to perform the method, an array, a method for identifying an agent
 CC that increases or decreases the expression of the polynucleotide sequence
 CC that is differentially expressed in neuronal tissue of a first animal
 CC subjected to pain, a method for identifying a compound which regulates
 CC the expression of a polynucleotide sequence which is differentially
 CC expressed in an animal subjected to pain, a method for identifying a
 CC compound that regulates the activity of one or more of the
 CC polynucleotides, a method for producing a pharmaceutical composition, a
 CC method for identifying a compound or small molecule that regulates the
 CC activity in an animal of one or more of the polypeptides given in the
 CC specification, a method for identifying a compound useful in treating
 CC pain and a pharmaceutical composition comprising the one or more
 CC polypeptides or their antibodies. The polynucleotide or the compound that
 CC modulates its activity is useful for preparing a medicament for treating
 CC pain (e.g. spinal segmental nerve injury (Chung), chronic constriction
 CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
 CC therapy). The sequence presented is a rat protein (shown in Table 2 of
 CC the specification) which is differentially expressed during pain. Note:
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic form directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.

CC Sequence 2531 AA;
 XX
 XX

Query Match 61.2%; Score 85; DB 7; Length 2531;
 Best Local Similarity 45.2%; Pred. No. 0.017;
 Matches 14; Conservative 2; Mismatches 15; Indels 0; Gaps 0;

QY 1 CDXXXCXXKXGNGCXDXXCNNNAACXXDXDC 31
 Db 1449 CELPECQDAGNKKVCNLCQNNHACGWDGDC 1479

RESULT 7
 ADE63709
 ID ADE63709 standard; protein; 2531 AA.
 XX
 AC ADE63709;
 XX
 DT 29-JAN-2004 (first entry)
 XX
 DE Rat Protein CAA40667, SEQ ID NO 9653.
 XX
 KW Rat, pain; neuronal tissue; gene therapy; spinal segmental nerve injury;
 KW chronic constriction injury; CCI; spared nerve injury; SNI; Chung.
 XX
 OS Rattus norvegicus.
 XX
 PN MO2003016475-A2.
 XX
 PD 27-FEB-2003.
 XX

PF 14-AUG-2002; 2002WO-US025765.
 XX
 PR 14-AUG-2001; 2001US-0312147P.
 PR 01-NOV-2001; 2001US-0346382P.
 PR 26-NOV-2001; 2001US-0333347P.
 XX
 PA (GENO) GEN HOSPITAL CORP.
 PA (FARB) BAYER AG.
 PI Woolf C, D'urso D, Befort K, Costigan M;
 DR WPI; 2003-268312/26.
 DR GENBANK; CAA40667.
 XX
 PT New composition comprising two or more isolated polypeptides, useful for
 PT preparing a medicament for treating pain in an animal.
 XX
 PS Claim 1; Page; 1017pp; English.
 CC The invention discloses a composition comprising two or more isolated rat
 CC or human polynucleotides or a polynucleotide which represents a fragment,
 CC derivative or allelic variation of the nucleic acid sequence. Also
 CC claimed are a vector comprising the novel polynucleotide, a host cell
 CC comprising the vector, a method for identifying a nucleotide sequence
 CC which is differentially regulated in an animal subjected to pain and a
 CC kit to perform the method, an array, a method for identifying an agent
 CC that increases or decreases the expression of the polynucleotide sequence
 CC that is differentially expressed in neuronal tissue of a first animal
 CC subjected to pain, a method for identifying a compound which regulates
 CC the expression of a polynucleotide sequence which is differentially
 CC expressed in an animal subjected to pain, a method for identifying a
 CC compound that regulates the activity of one or more of the
 CC polynucleotides, a method for producing a pharmaceutical composition, a
 CC method for identifying a compound or small molecule that regulates the
 CC activity in an animal of one or more of the polypeptides given in the
 CC specification, a method for identifying a compound useful in treating
 CC pain and a pharmaceutical composition comprising the one or more
 CC polypeptides or their antibodies. The polynucleotide or the compound that
 CC modulates its activity is useful for preparing a medicament for treating
 CC pain (e.g. spinal segmental nerve injury (Chung), chronic constriction
 CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
 CC therapy). The sequence presented is a rat protein (shown in Table 2 of
 CC the specification) which is differentially expressed during pain. Note:
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic form directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 2531 AA;
 Query Match 61.2%; Score 85; DB 7; Length 2531;
 Best Local Similarity 45.2%; Pred. No. 0.017;
 Matches 14; Conservative 2; Mismatches 15; Indels 0; Gaps 0;
 QY 1 CDXXXCXKXGNGCXCDXCNNAACXXDXGDC 31
 Db 1449 CELPECOEDAGNKCVCNLQCNNAHACGWDGDC 1479
 RESULT 8
 ADB63701 ID ADB63701 standard; protein; 2531 AA.
 XX
 AC ADE63701;
 XX
 DT 29-JAN-2004 (first entry)
 XX
 DE Rat Protein CAA40667, SEQ ID NO 9645.
 XX
 XX Rat; pain; neuronal tissue; gene therapy; spinal segmental nerve injury;
 KM chronic constriction injury; CCI, spared nerve injury; SNI; Chung.
 XX
 OS Rattus norvegicus.
 XX

EN WO2003016475-A2.
 XX
 PD 27-FEB-2003.
 XX
 XX
 PF 14-AUG-2002; 2002WO-US025765.
 XX
 PR 14-AUG-2001; 2001US-0312147P.
 PR 01-NOV-2001; 2001US-0346382P.
 PR 26-NOV-2001; 2001US-0333347P.
 XX
 PA (GENO) GEN HOSPITAL CORP.
 PA (FARB) BAYER AG.
 PI Woolf C, D'urso D, Befort K, Costigan M;
 DR WPI; 2003-268312/26.
 DR GENBANK; CAA40667.
 XX
 PT New composition comprising two or more isolated polypeptides, useful for
 PT preparing a medicament for treating pain in an animal.
 XX
 PS Claim 1; Page; 1017pp; English.
 CC The invention discloses a composition comprising two or more isolated rat
 CC or human polynucleotides or a polynucleotide which represents a fragment,
 CC derivative or allelic variation of the nucleic acid sequence. Also
 CC claimed are a vector comprising the novel polynucleotide, a host cell
 CC comprising the vector, a method for identifying a nucleotide sequence
 CC which is differentially regulated in an animal subjected to pain and a
 CC kit to perform the method, an array, a method for identifying an agent
 CC that increases or decreases the expression of the polynucleotide sequence
 CC that is differentially expressed in neuronal tissue of a first animal
 CC subjected to pain, a method for identifying a compound which regulates
 CC the expression of a polynucleotide sequence which is differentially
 CC expressed in an animal subjected to pain, a method for identifying a
 CC compound that regulates the activity of one or more of the
 CC polynucleotides, a method for producing a pharmaceutical composition, a
 CC method for identifying a compound or small molecule that regulates the
 CC activity in an animal of one or more of the polypeptides given in the
 CC specification, a method for identifying a compound useful in treating
 CC pain and a pharmaceutical composition comprising the one or more
 CC polypeptides or their antibodies. The polynucleotide or the compound that
 CC modulates its activity is useful for preparing a medicament for treating
 CC pain (e.g. spinal segmental nerve injury (Chung), chronic constriction
 CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
 CC therapy). The sequence presented is a rat protein (shown in Table 2 of
 CC the specification) which is differentially expressed during pain. Note:
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic form directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 2531 AA;
 Query Match 61.2%; Score 85; DB 7; Length 2531;
 Best Local Similarity 45.2%; Pred. No. 0.017;
 Matches 14; Conservative 2; Mismatches 15; Indels 0; Gaps 0;
 QY 1 CDXXXCXKXGNGCXCDXCNNAACXXDXGDC 31
 Db 1449 CELPECOEDAGNKCVCNLQCNNAHACGWDGDC 1479
 RESULT 9
 ABB07821 ID ABB07821 standard; protein; 2444 AA.
 XX
 AC ABB07821;
 XX
 DT 03-JUL-2002 (first entry)
 XX
 DE Constitutively active notch-1 protein.
 XX
 KM Cell differentiation; notch; epidermis; cytostatic; dermatological;
 XX

CC diseases, immunodeficiency, HIV (human immuno deficiency virus), viral,
CC bacterial and fungal infections, hepatitis, rheumatoid arthritis, asthma,
CC haemolytic, thrombolytic, haemostatic diseases, thrombocytopenia,
CC ulcers, Alzheimer's disease, Parkinson's disease, Huntington's disease,
CC ocular disease, muscular diseases, growth disorders, loss of libido,
CC stress, depression, pain and epilepsy. They are useful for preventing
CC chemotherapy side effects and as contraceptives. Sequences of the
CC invention are also useful for gene therapy. The present sequence is human
CC Notch-like protein, MO1a
XX
SQ Sequence 2469 AA;

Query Match 60.4%; Score 84; DB 5; Length 2469;
Best Local Similarity 45.2%; Pred. No. 0.022;
Matches 14; Conservative 2; Mismatches 15; Indels 0; Gaps 0;

QY 1 CDXXXCXKXGKCDXXCNNAACXXDGXDC 31
Db 1423 CLSQYCADRKARDGVCDEACNSHACQWDGDC 1453

RESULT 11

ADD18192
ID ADD18192 standard; protein; 2469 AA.

AC ADD18192;

DT 15-JAN-2004 (first entry)

DE Human molecule (MOL) protein MO1a.

XX molecule protein; MOL protein; MOLX; MOLX agonist; MOLX antagonist;

KM cardiant; antidiabetic; antiarteriosclerotic; gene therapy;

KW MOLX-associated disorder; cardiomyopathy; diabetes; atherosclerosis;

KM human; MO1a.

XX Homo sapiens.

OS WO2003003984-A2.

PN 16-JAN-2003.

XX 03-JUL-2002; 2002WO-US021268.

XX 05-JUL-2001; 2001US-0303168P.

XX 05-JUL-2001; 2001US-0303241P.

XX 26-SEP-2001; 2001US-00965212.

XX 26-SEP-2001; 2001US-0096545.

XX 26-SEP-2001; 2001US-0368996P.

XX 01-APR-2002; 2002US-0369065P.

XX 08-MAY-2002; 2002US-0378730P.

XX 30-MAY-2002; 2002US-0384327P.

XX 07-JUN-2002; 2002US-0386816P.

XX 17-JUN-2002; 2002US-00174372.

XX (CURA-) CURAGEN CORP.

XX PI Fernandes ER, Vernet CAM, Shinkels RA, Anderson DW, Padigaru M,
XX PI Boldog FU, Li L, Shenoy SG, Casman SU, Rastelli L, Alsobrook JP,
XX PI Burgees CE, Grose WM, Gusev VY, Ji W, Lepley DW, Liu X, Mezick AJ,
XX PI Patturajan M, Shen L, Spaderna SK, Spytek KA, Szekeres ES,
XX PI Tautner RJ, Tchernev VT, Zerhusen BD, Voss EZ,
XX
XX WPI; 2003-210304/20.
XX N-PSDB; ADD18191.

XX New MOLX polypeptide, nucleic acid or MOLX-specific antibody, useful for
XX PT preparing a composition for treating or preventing a MOLX-associated
XX PT disorder, e.g., cardiomyopathy, diabetes or atherosclerosis.
XX
XX Claim 1; SEQ ID NO 2; 371pp; English.

CC This invention relates to novel human nucleic acid sequences which encode
CC novel molecule (MOL) proteins numbered MO1-23, referred to generally in
CC the specification as MOLX. Compounds which modulate the function of the
CC MOLX proteins of the invention, MOLX agonists or antagonists, may have
CC cardiant, antidiabetic or antiarteriosclerotic activities. In addition,
CC the DNA and protein sequences disclosed may prove useful for gene
CC therapy. The protein, nucleic acid or antibody is useful for preparing a
CC composition for treating or preventing a MOLX-associated disorder, for
CC example cardiomyopathy, diabetes or atherosclerosis. The present sequence
CC is the amino acid sequence of a MOL protein of the invention.
XX
SQ Sequence 2469 AA;

Query Match 60.4%; Score 84; DB 7; Length 2469;
Best Local Similarity 45.2%; Pred. No. 0.022;
Matches 14; Conservative 2; Mismatches 15; Indels 0; Gaps 0;

QY 1 CDXXXCXKXGKCDXXCNNAACXXDGXDC 31
Db 1423 CLSQYCADRKARDGVCDEACNSHACQWDGDC 1453

RESULT 12

AA027065
ID AA027065 standard; protein; 2471 AA.

XX AA027065;

DT 22-MAY-2003 (first entry)

DE Human Notch protein homologue sequence, SEQ ID NO 19.

XX Notch protein; cell fat disorder; differentiation; cancer; breast; colon;

KW cervical cancer; pre-neoplastic; non-malignant state; neoplastic;

KW malignant; human.

XX Homo sapiens.

OS WO9407474-A1.

PN 14-APR-1994.

XX 30-SEP-1993; 93WO-US009338.

XX 30-SEP-1992; 92US-00955012.

XX 25-JUN-1993; 93US-00083590.

XX (UYIA) UNIT YALE.

XX Artavanis-Tsakonas S, Fehon RG, Zagouras P, Blumweller CM;
XX PI WPI; 1994-135180/16.

XX Notch protein and nuclear acid compositions - is used for treatment of
XX PT disorders of cell fate or differentiation esp. breast, colon or cervical
XX PT cancer.
XX
XX Claim 78; Fig 13; 233pp; English.

XX The invention relates to novel pharmaceutical compositions comprising a
XX CC therapeutically effective amount of a Notch protein and a
XX CC pharmaceutically acceptable carrier. The compositions can be used for
XX CC treatment of disorders of cell fate or differentiation. The therapeutic
XX CC compositions include Notch proteins and analogues, derivatives and
XX CC fragments, antibodies, and the nucleic acids encoding the Notch proteins.
XX CC The compositions are preferably administered to a cancerous condition,
XX CC e.g. breast, colon or cervical cancer, or to prevent progression from a
XX CC pre-neoplastic or non-malignant state into a neoplastic or malignant
XX CC state. This sequence represents a human Notch protein homologue of the
XX CC invention
XX
XX Sequence 2471 AA;

Query Match 60.4%; Score 84; DB 2; Length 2471;
 Best Local Similarity 45.2%; Pred. No. 0.022;
 Matches 14; Conservative 2; Mismatches 15; Indels 0; Gaps 0;

1 CDXXXXXXKXGKXCDXXCXXNNAACXXDXDC 31
 1425 CLSGYCADKARDGVCDACNSHACQMDGDC 1455

RESULT 13
 ID AAY06816 standard; protein; 2471 AA.
 AC AAY06816;
 DT 05-JUL-1999 (first entry)
 DE Human Notch2 (humN2) protein sequence.
 KW Notch; Ncm; epidermal growth factor; EGF; repeat domain; cancer;
 KW transmembrane domain; cell-face disorder; proliferative disease;
 KW signal transduction; human; Notch2; humN2.
 OS Homo sapiens.
 XX

Key Location/Qualifiers
 FT Cleavage-site 25 /note= "potential signal cleavage site"
 FT Region 26..1413 /note= "EGF repeat region"
 FT Region 1425..1537 /note= "Lin-12/notch repeats"
 FT Domain 1678..1699 /note= "transmembrane domain"
 FT Region 1822..2035 /note= "Ankyrin repeats"
 FT Region 2130..2133 /note= "nuclear localisation signal (NLS) sequence"
 FT Modified-site 2149..2152 /note= "putative CKII phosphorylation site"
 FT Modified-site 2156..2159 /note= "putative cdc2 phosphorylation site"
 FT Modified-site 2166..2169 /note= "putative cdc2 phosphorylation site"
 FT Region 2172..2174 /note= "nuclear localisation signal (BNTS) sequence"
 FT Region 2385..2445 /note= "PST-containing region"
 XX

PN WO9904746-A2.
 XX
 PD 04-FEB-1999.
 XX
 PF 23-JUL-1998; 98WO-US015333.
 XX
 PR 23-JUL-1997; 97US-00899232.
 XX
 PA (UYVA) UNITV YALB.
 XX
 PI Artavanis-Tsakonas S, Qi H, Rand MD;
 XX
 DR WPI; 1999-153290/13.
 XX
 PT Measuring and detecting activation of Notch protein in cells from cell
 PT surface expression - useful for, e.g. identifying modulators of
 PT activation, potentially useful for treating cancer.
 XX
 PS Disclosure; Fig 2A-D; 94pp; English.
 XX
 CC The invention relates to methods of measuring activation of Notch in a
 CC cell. Activated notch can be detected from: (i) expression of: (a) Notch
 CC on the cell surface, or (b) one or both of the Notch cleavage products
 CC Nec and Ncm; (ii) presence of Notch fragments, i.e. one or both of an N-

CC terminal fragment (F1) terminating between the epidermal growth factor -
 CC like (EGF) repeat domain and the transmembrane domain or a C-terminal
 CC fragment (F2) that has its N-terminus between the specified domains, or
 CC fragments of molecular weights about 270, 200, 170, 140, 110, 100, 90 and
 CC 85 kDa; or (iii) presence of a Notch heterodimer containing a linkage
 CC sensitive to reducing agents. The method is used to study or manipulate
 CC differentiation processes and to screen/diagnose cell-face disorders
 CC (particularly cancer or other proliferative diseases involving abnormal
 CC Notch activation). Modulators of Notch activation are useful for studying
 CC or manipulating differentiation and possibly for treating cancer. The F1-
 CC F2 heterodimer is the active form of Notch that mediates signal
 CC transduction and binds ligands such as Delta and Serrate. F1 and F2 are
 CC generated by proteolysis of full-length Notch in the trans-Golgi. The
 CC present sequence represents the human Notch2 (humN2) sequence
 XX

Sequence 2471 AA;
 SQ

Query Match 60.4%; Score 84; DB 2; Length 2471;
 Best Local Similarity 45.2%; Pred. No. 0.022;
 Matches 14; Conservative 2; Mismatches 15; Indels 0; Gaps 0;

1 CDXXXXXXKXGKXCDXXCXXNNAACXXDXDC 31
 1425 CLSGYCADKARDGVCDACNSHACQMDGDC 1455

RESULT 14
 ID AAG79774 standard; protein; 2471 AA.
 AC AAG79774;
 DT 01-APR-2003 (first entry)
 DE Human Notch 2.
 KW Penetratin; antenapedia; homeodomain; intracellular delivery; anaemia;
 KW transport protein; Notch; signalling modulation; rheumatoid arthritis;
 KW T-cell mediated disease; asthma; allergy; graft rejection; autoimmunity;
 KW tumour; Plasmodium; infection; HIV; pseudomonas; hepatitis C; measles;
 KW multiple sclerosis; diabetes; progressive bulbar palsy; neurodegeneration;
 KW organ; transplantation; bone marrow; cell fate; small cell lung cancer;
 KW cancer; kidney; uterus; prostate; bladder; ovary; colon; breast;
 KW nervous system; lesion; traumatic; ischaemic; malignant; infectious; HIV;
 KW degenerative; demyelinated; diabetes; Parkinson's disease;
 KW Charcot-Marie-Tooth disease; liver cirrhosis; Alzheimer's disease;
 KW hypertrophic scar formation; psoriasis; organ regeneration;
 KW tissue engineering.
 XX
 OS Homo sapiens.
 XX
 PD WO200296952-A2.
 XX
 PF 24-MAY-2002; 2002WO-GB002438.
 XX
 PR 25-MAY-2001; 2001GB-00012818.
 XX
 PA (LORA-) LORANTIS LTD.
 XX
 PI Solari RCE, Champion BR, Ward GA;
 XX
 DR WPI; 2003-156790/15.
 XX
 PT Conjugate for delivering a protein for Notch signal modulation into cells
 PT in vitro/in vivo, and in treating T-cell mediated disease, comprises a
 PT transport protein sequence and sequence for Notch signaling modulation.
 XX
 PS Disclosure; Fig 6; 80pp; English.
 XX
 CC The sequences given in AAG79773-74 represent human Notch 1 and Notch 2
 CC which may be used in the conjugate of the invention. The conjugate

comprises first and second sequences, where the first sequence comprises a transport protein or a polynucleotide coding for a transport protein and the second sequence comprises a polypeptide or polynucleotide for Notch signalling modulation. The conjugate is useful for transporting a protein for Notch signalling modulation or a polynucleotide sequence encoding it into a cell. Pharmaceuticals comprising the conjugate of the invention are useful in the treatment of a T-cell mediated disease and for the preparation of a medicament for the preventing and/or treatment of disease or infection, especially T-cell mediated disease. T-cell mediated diseases include asthma, allergy, graft rejection, autoimmunity, tumour induced aberrations to the T cell system and infectious diseases such as caused by Plasmodium sp., HIV, pseudomonas, hepatitis C, measles, multiple sclerosis, rheumatoid arthritis, diabetes, organ transplantation or bone marrow transplantation. The conjugate is also useful for altering the fate of a cell, tissue or organ type by altering Notch pathway function in the cell, in the treatment of malignant and pre-neoplastic disorders e.g. small cell lung cancer and cancers of the kidney, uterus, prostate, bladder, ovary, colon and breast. It is further useful for treating nervous system disorders including neurological lesions including traumatic lesions resulting from physical injuries, ischaemic lesions, malignant lesions, infectious lesions caused by HIV, degenerative lesions and diseases and demyelinated lesions. Diseases treatable include diabetes, Parkinson's, Alzheimer's disease, progressive bulbar palsy and Charcot-Marie-Tooth disease. The conjugate is also useful for promoting tissue regeneration and repair, treating diseases associated with defective tissue repair and regeneration for e.g. cirrhosis of the liver, hypertrophic scar formation and psoriasis, and also for treating neuropenia, or anaemia and in techniques of organ regeneration and tissue engineering. The conjugate has improved efficacy compared to conventional treatments, improved cellular uptake of the therapeutic agent, improved water solubility, reduction of side effects and cellular bioavailability and decreased occurrence of drug resistance

Sequence 2471 AA:

Query Match 60.4%; Score 84; DB 6; Length 2471;
Best Local Similarity 45.2%; Pred. No. 0.022;
Matches 14; Conservative 2; Mismatches 15; Indels 0; Gaps 0;

QY 1 CDXXXCXKXGNGCXNNACXXDXDC 31
DB 1425 CLSQYCADRKADGVCDACNSHACQWDGDC 1455

RESULT 15

ABP72572
ID ABP72572 standard; protein: 2471 AA.

AC ABP72572;

DT 29-MAY-2003 (first entry)

DE Human Notch 2.

Human; Notch 2; signal transduction; modulator; cytostatic;
antiasthmatic; antiallergic; antirheumatic; antiarthritic; antidiabetic;
antithyroid; neuroprotective; ophthalmological; antiinflammatory;
hepatotropic; dermatological; immunosuppressive; antibacterial; virucide;
antiarteriosclerotic; cardiant; antiarteriosclerotic; vasotropic;
antidiacer; nephrotropic; antiparkinsonian; nootropic; gynaecological;
protozoacide.

XX Homo sapiens.

EN WO2003012441-A1.

XX 13-FEB-2003.

PF 25-JUL-2002; 2002WO-GB003397.

PR 25-JUL-2001; 2001GB-00018153.

PR 05-APR-2002; 2002GB-00007930.

PR 28-MAY-2002; 2002GB-00012282.

PR 28-MAY-2002; 2002GB-00012283.

XX (LORA-) LORANTIS LTD.

XX Bodmer MW, Briend ECP, Champion BR, McKenzie GJ, Tugal T;

XX Ward GA, Young LL;

XX WPI; 2003-248208/24.

PT Detecting modulators of notch or immune signalling useful for treating cancer, asthma, allergy, lupus erythematosus, diabetes, by monitoring Notch signalling in a cell of the immune system in the presence of modulator.

XX Disclosure; Fig 36; 184pp; English.

The present sequence is the protein sequence of human Notch 2. The invention provides a method for detecting modulators of Notch signalling. This involves monitoring the effect of a candidate modulator on Notch signalling in a cell of the immune system, such as a T cell, a B cell or an antigen presenting cell, e.g. by monitoring the level of expression of a target gene such as an endogenous target gene of the Notch signalling pathway or a reporter gene. The modulator may comprise a Notch intracellular domain (see also ABP72552) or a polynucleotide encoding it. The method detects modulators of Notch signalling by monitoring the effect of a candidate modulator on Notch signalling in a cell of the immune system, such as a T cell, a B cell or an antigen presenting cell. The method is useful for preparing a medicament for treating a disease or condition of, or related to, the immune system, such as T-cell, B-cell or APC mediated disease (all claimed), including cancer, acute and chronic immune and autoimmune pathologies, infectious diseases, inflammatory diseases, neurodegenerative diseases, alcohol-induced hepatitis, other diseases related to angiogenesis, cardiovascular conditions and pulmonary diseases

Sequence 2471 AA:

Query Match 60.4%; Score 84; DB 6; Length 2471;
Best Local Similarity 45.2%; Pred. No. 0.022;
Matches 14; Conservative 2; Mismatches 15; Indels 0; Gaps 0;

QY 1 CDXXXCXKXGNGCXNNACXXDXDC 31
DB 1425 CLSQYCADRKADGVCDACNSHACQWDGDC 1455

Search completed: May 19, 2004, 15:57:48
Job time : 5.076 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 19, 2004, 15:55:54 ; Search time 1.29658 Seconds
(without alignments)
1234.330 Million cell updates/sec

Title: US-09-903-199-4

Perfect score: 139

Sequence: 1 CDXXXCXKXGXGCDXXCNMAA CXXDXDC 31

Scoring table: BLOSUM62

Gap 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patente AA: *
1: /cgn2_6/ptodata/2/1aa/5A COMB.pep: *
2: /cgn2_6/ptodata/2/1aa/5B COMB.pep: *
3: /cgn2_6/ptodata/2/1aa/6A COMB.pep: *
4: /cgn2_6/ptodata/2/1aa/6B COMB.pep: *
5: /cgn2_6/ptodata/2/1aa/PCITUS COMB.pep: *
6: /cgn2_6/ptodata/2/1aa/backfile1.pep: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	105	75.5	1139	1 US-08-537-210A-4	Sequence 4, Appli
2	105	75.5	1139	1 US-08-113-825-4	Sequence 4, Appli
3	105	75.5	2703	1 US-08-185-432-19	Sequence 19, Appli
4	105	75.5	2703	4 US-08-899-232-4	Sequence 4, Appli
5	84	60.4	1015	1 US-08-537-210A-1	Sequence 1, Appli
6	84	60.4	1015	3 US-08-113-825-1	Sequence 2, Appli
7	84	60.4	1068	1 US-08-537-210A-2	Sequence 2, Appli
8	84	60.4	1068	3 US-09-113-825-2	Sequence 2, Appli
9	84	60.4	2471	1 US-08-185-432-16	Sequence 16, Appli
10	84	60.4	2471	1 US-08-083-590A-19	Sequence 19, Appli
11	84	60.4	2471	3 US-08-532-384-19	Sequence 19, Appli
12	84	60.4	2471	4 US-08-899-232-1	Sequence 17, Appli
13	84	60.4	2556	1 US-08-185-432-17	Sequence 20, Appli
14	84	60.4	2556	1 US-08-083-580A-20	Sequence 20, Appli
15	84	60.4	2556	3 US-08-532-384-20	Sequence 20, Appli
16	84	60.4	2556	4 US-08-899-232-2	Sequence 2, Appli
17	83	59.7	2321	4 US-09-230-652-2	Sequence 2, Appli
18	80	57.6	1064	1 US-08-537-210A-3	Sequence 3, Appli
19	80	57.6	1064	3 US-09-113-825-3	Sequence 3, Appli
20	80	57.6	2523	1 US-08-185-432-18	Sequence 16, Appli
21	80	57.6	2523	4 US-08-899-232-3	Sequence 3, Appli
22	79	56.8	1964	4 US-09-467-997-1	Sequence 1, Appli
23	78	56.1	928	4 US-09-635-872A-1	Sequence 1, Appli
24	78	56.1	928	4 US-09-636-077A-1	Sequence 1, Appli
25	78	56.1	928	4 US-09-636-060C-1	Sequence 1, Appli
26	78	56.1	928	4 US-09-986-552-1	Sequence 1, Appli
27	77	55.4	1078	1 US-08-264-534-32	Sequence 32, Appli

28	77	55.4	1078	1 US-08-083-590A-11	Sequence 11, Appli
29	77	55.4	1078	2 US-08-465-500-32	Sequence 32, Appli
30	77	55.4	1078	1 US-08-346-128-32	Sequence 32, Appli
31	77	55.4	1078	3 US-08-532-384-11	Sequence 11, Appli
32	77	55.4	1078	3 US-08-893-828-32	Sequence 32, Appli
33	73	52.5	502	4 US-09-635-872A-13	Sequence 13, Appli
34	73	52.5	502	4 US-09-636-077A-13	Sequence 13, Appli
35	73	52.5	502	4 US-09-636-060C-13	Sequence 13, Appli
36	73	52.5	502	4 US-09-986-552-13	Sequence 13, Appli
37	64	46.0	1791	4 US-09-827-998-3	Sequence 3, Appli
38	62	44.6	908	4 US-09-635-872A-15	Sequence 15, Appli
39	62	44.6	908	4 US-09-636-077A-15	Sequence 15, Appli
40	62	44.6	908	4 US-09-636-060C-15	Sequence 15, Appli
41	62	44.6	908	4 US-09-986-552-15	Sequence 15, Appli
42	59	42.4	1385	4 US-09-827-998-16	Sequence 16, Appli
43	57	41.0	1770	4 US-09-827-998-10	Sequence 10, Appli
44	48.5	34.9	233	4 US-09-216-393B-110	Sequence 10, Appli
45	48.5	34.9	616	4 US-09-608-790-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-08-537-210A-4
; Sequence 4, Application US/08537210A
; Patent No. 5780300
; GENERAL INFORMATION:
; APPLICANT: Artavanis-Tsakonas, Spyridon
; APPLICANT: Fortini, Mark
; APPLICANT: Matsuno, Kenji
; TITLE OF INVENTION: MANIPULATION OF NON-TERMINALLY
; TITLE OF INVENTION: DIFFERENTIATED CELLS USING THE NOTCH PATHWAY
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036/2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/537,210A
; FILING DATE: 29-SEP-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Mierock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7326-027
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-790-9090
; TELEFAX: 212-869-8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1139 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; FEATURE:
; NAME/KEY: Dros N
; LOCATION: 1189...2337
; OTHER INFORMATION: Highly conserved ankyrin repeat
; OTHER INFORMATION: region of No. 57803000ch
US-08-537-210A-4
Query Match 75.5%, Score 105; DB 1; Length 1139;

Best Local Similarity 54.8%; Pred. No. 3,1e-06;
Matches 17; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 1 CDXXXCKXKXGNGKCDXXCNNAACXXDGXDC 31
DB 294 CDKRGCTEKQNGICDSDCNTYACNFDGNDG 324

RESULT 2

US-09-113-825-4
Sequence 4, Application US/09113825
Patent No. 6149902

GENERAL INFORMATION:

APPLICANT: Artavanis-Tsakonas, Spyridon
APPLICANT: Fortini, Mark
APPLICANT: Matsuno, Kenji
TITLE OF INVENTION: MANIPULATION OF NON-TERMINALLY
TITLE OF INVENTION: DIFFERENTIATED CELLS USING THE NOTCH PATHWAY
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10036/2711

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS

SOFTWARE: FastSeq Version 2.0

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/113,825
FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/537,210
FILING DATE: 29-SEP-1995

ATTORNEY/AGENT INFORMATION:

NAME: Mistrock, S. Leslie
REGISTRATION NUMBER: 18,872

REFERENCE/DOCKET NUMBER: 7326-027
TELECOMMUNICATION INFORMATION:

TELEPHONE: 212-790-9090
TELEFAX: 212-869-8864

TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 1139 amino acids
TYPE: amino acid

STRANDEDNESS:

TOPOLOGY: unknown
MOLECULE TYPE: protein

FEATURE:

NAME/KEY: Dros N
LOCATION: 1189...2327

OTHER INFORMATION: Highly conserved ankyrin repeat
OTHER INFORMATION: region of No. 6149902ch

US-09-113-825-4

Query Match 75.5%; Score 105; DB 3; Length 1139;
Best Local Similarity 54.8%; Pred. No. 3,1e-06;
Matches 17; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 1 CDXXXCKXKXGNGKCDXXCNNAACXXDGXDC 31
DB 294 CDKRGCTEKQNGICDSDCNTYACNFDGNDG 324

RESULT 3

US-08-185-432-19
Sequence 19, Application US/08185432
Patent No. 5750652

GENERAL INFORMATION:

APPLICANT: Artavanis-Tsakonas, Spyridon
APPLICANT: Buseau, Isabelle
APPLICANT: Diederich, Robert J.

APPLICANT: Xu, Tian
APPLICANT: Matsuno, Kenji

TITLE OF INVENTION: DUTEX PROTEINS, NUCLEIC ACIDS, AND
TITLE OF INVENTION: ANTIBODIES, AND RELATED METHODS AND COMPOSITIONS
NUMBER OF SEQUENCES: 23

CORRESPONDENCE ADDRESS:

ADDRESSEE: PENNIE & EDMONDS
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.

ZIP: 10036-2711

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/185,432
FILING DATE: 21-JAN-1994

CLASSIFICATION: 530

ATTORNEY/AGENT INFORMATION:

NAME: Mistrock, S. Leslie
REGISTRATION NUMBER: 18,872

REFERENCE/DOCKET NUMBER: 7326-006
TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741

TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 19:

SEQUENCE CHARACTERISTICS:

LENGTH: 2703 amino acids
TYPE: amino acid

TOPOLOGY: unknown

MOLECULE TYPE: protein

US-08-185-432-19
Query Match 75.5%; Score 105; DB 1; Length 2703;
Best Local Similarity 54.8%; Pred. No. 6,3e-06;
Matches 17; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 1 CDXXXCKXKXGNGKCDXXCNNAACXXDGXDC 31
DB 1482 CDKRGCTEKQNGICDSDCNTYACNFDGNDG 1512

US-08-899-232-4
Sequence 4, Application US/08899232
Patent No. 6436650

GENERAL INFORMATION:

APPLICANT: Artavanis-Tsakonas, Spyridon
APPLICANT: Qi, Hulin

TITLE OF INVENTION: ACTIVATED FORMS OF NOTCH AND METHODS BASED THEREON
FILE REFERENCE: 7326-046

CURRENT APPLICATION NUMBER: US/08/899,232
CURRENT FILING DATE: 1997-07-23

NUMBER OF SEQ ID NOS: 4
SOFTWARE: Patentin Ver. 2.0

SEQ ID NO 4

LENGTH: 2703

TYPE: PRT

ORGANISM: Drosophila sp.

US-08-899-232-4

Query Match 75.5%; Score 105; DB 4; Length 2703;
Best Local Similarity 54.8%; Pred. No. 6,3e-06;
Matches 17; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

OY 1 CDXXXCKXXGNGXCDXXCNNAACXXDXDC 31
DB 1482 CDRGCTEKXGNGICDSDCNTYACNPDGND 1512

RESULT 5

US-08-537-210A-1
Sequence 1, Application US/08537210A
Patent No. 5780300
GENERAL INFORMATION:
APPLICANT: Artavanis-Tsakonas, Spyridon
APPLICANT: Fortini, Mark
APPLICANT: Matsuno, Kenji
TITLE OF INVENTION: MANIPULATION OF NON-TERMINALLY
TITLE OF INVENTION: DIFFERENTIATED CELLS USING THE NOTCH PATHWAY
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10036/2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/537,210A
FILING DATE: 29-SEP-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Mastrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 7326-027
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-790-9090
TELEFAX: 212-869-8864
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1015 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: unknown
MOLECULE TYPE: protein
FEATURE:
NAME/KEY: hum N (Human No. 5780300ch 2)
LOCATION: 1155...2169
OTHER INFORMATION: Highly conserved ankyrin repeat
OTHER INFORMATION: region of No. 5780300ch
US-08-537-210A-1
Query Match 60.4%; Score 84; DB 1; Length 1015;
Best Local Similarity 45.2%; Pred. No. 0.0014;
Matches 14; Conservative 2; Mismatches 15; Indels 0; Gaps 0;
OY 1 CDXXXCKXXGNGXCDXXCNNAACXXDXDC 31
DB 271 CLSQYCADKARDGVCDKCNHACQWDGDC 301

RESULT 6

US-09-113-825-1
Sequence 1, Application US/09113825
Patent No. 6149902
GENERAL INFORMATION:
APPLICANT: Artavanis-Tsakonas, Spyridon
APPLICANT: Fortini, Mark
APPLICANT: Matsuno, Kenji
TITLE OF INVENTION: MANIPULATION OF NON-TERMINALLY
TITLE OF INVENTION: DIFFERENTIATED CELLS USING THE NOTCH PATHWAY

NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10036/2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/113,825
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/537,210
FILING DATE: 29-SEP-1995
ATTORNEY/AGENT INFORMATION:
NAME: Mastrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 7326-027
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-790-9090
TELEFAX: 212-869-8864
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1015 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: unknown
MOLECULE TYPE: protein
FEATURE:
NAME/KEY: hum N (Human No. 6149902ch 2)
LOCATION: 1155...2169
OTHER INFORMATION: Highly conserved ankyrin repeat
OTHER INFORMATION: region of No. 6149902ch
US-09-113-825-1

Query Match 60.4%; Score 84; DB 3; Length 1015;
Best Local Similarity 45.2%; Pred. No. 0.0014;
Matches 14; Conservative 2; Mismatches 15; Indels 0; Gaps 0;

OY 1 CDXXXCKXXGNGXCDXXCNNAACXXDXDC 31
DB 271 CLSQYCADKARDGVCDKCNHACQWDGDC 301

RESULT 7

US-08-537-210A-2
Sequence 2, Application US/08537210A
Patent No. 5780300
GENERAL INFORMATION:
APPLICANT: Artavanis-Tsakonas, Spyridon
APPLICANT: Fortini, Mark
APPLICANT: Matsuno, Kenji
TITLE OF INVENTION: MANIPULATION OF NON-TERMINALLY
TITLE OF INVENTION: DIFFERENTIATED CELLS USING THE NOTCH PATHWAY
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10036/2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS

SOFTWARE: FastSeq Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/537,210A
FILING DATE: 29-SEP-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Mistrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 7326-027
TELEPHONE: 212-790-9090
TELEFAX: 212-869-8864
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1068 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: unknown
MOLECULE TYPE: protein
FEATURE:
NAME/KEY: Human NI (TAN-1)
LOCATION: 1152...2219
OTHER INFORMATION: Highly conserved ankyrin repeat
OTHER INFORMATION: region of No. 5780300ch
US-08-537-210A-2

Query Match 60.4%; Score 84; DB 1; Length 1068;
Best Local Similarity 45.2%; Pred. No. 0.0015;
Matches 14; Conservative 1; Mismatches 16; Indels 0; Gaps 0;

QY 1 CDXXXCXXKXGNGXCDXXCNNAACXXDXDC 31
DB 299 CELPECGEDAGNKVCISLQCNHACGWDGDC 329

RESULT 8
US-09-113-825-2
Sequence 2, Application US/09113825
Patent No. 6199902
GENERAL INFORMATION:
APPLICANT: Artavanis-Tsakonas, Spyridon
APPLICANT: Fortini, Mark
APPLICANT: Matsuno, Kenji
TITLE OF INVENTION: MANIPULATION OF NON-TERMINALLY
TITLE OF INVENTION: DIFFERENTIATED CELLS USING THE NOTCH PATHWAY
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10036/2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/113,825
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/537,210
FILING DATE: 29-SEP-1995
ATTORNEY/AGENT INFORMATION:
NAME: Mistrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 7326-027
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-790-9090
TELEFAX: 212-869-8864

TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1068 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: unknown
MOLECULE TYPE: protein
FEATURE:
NAME/KEY: Human NI (TAN-1)
LOCATION: 1152...2219
OTHER INFORMATION: Highly conserved ankyrin repeat
OTHER INFORMATION: region of No. 6149902ch
US-09-113-825-2

Query Match 60.4%; Score 84; DB 3; Length 1068;
Best Local Similarity 45.2%; Pred. No. 0.0015;
Matches 14; Conservative 1; Mismatches 16; Indels 0; Gaps 0;

QY 1 CDXXXCXXKXGNGXCDXXCNNAACXXDXDC 31
DB 299 CELPECGEDAGNKVCISLQCNHACGWDGDC 329

RESULT 9
US-08-185-432-16
Sequence 16, Application US/08185432
Patent No. 5750652
GENERAL INFORMATION:
APPLICANT: Artavanis-Tsakonas, Spyridon
APPLICANT: Buseau, Isabelle
APPLICANT: Diederich, Robert J.
APPLICANT: Xu, Tian
APPLICANT: Matsuno, Kenji
TITLE OF INVENTION: DELTEX PROTEINS, NUCLEIC ACIDS, AND
TITLE OF INVENTION: ANTIBODIES, AND RELATED METHODS AND COMPOSITIONS
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: PENNIE & EDMONDS
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/185,432
FILING DATE: 21-JAN-1994
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Mistrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 7326-006
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 2471 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-185-432-16

Query Match 60.4%; Score 84; DB 1; Length 2471;
Best Local Similarity 45.2%; Pred. No. 0.003;
Matches 14; Conservative 2; Mismatches 15; Indels 0; Gaps 0;

QY 1 CDXXXXXXGNGXCDXXCNNAACXXDXGDC 31
 Db 1425 CLSQYCADKARDGVCDACNSHACQMDGDC 1455

RESULT 10
 US-08-083-590A-19
 ; Sequence 19, Application US/08083590A
 ; Patent No. 5786158
 ; GENERAL INFORMATION:
 ; APPLICANT: Artavanis-Tsakonas, S. et al.
 ; TITLE OF INVENTION: Therapeutic And Diagnostic Methods
 ; TITLE OF INVENTION: And Compositions Based On No. 5786158ch Proteins And
 ; TITLE OF INVENTION: Nucleic Acids
 ; NUMBER OF SEQUENCES: 21
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Pennie & Edmonds
 ; STREET: 1155 Avenue of the Americas
 ; CITY: New York
 ; STATE: New York
 ; COUNTRY: U.S.A.
 ; ZIP: 10036
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/083,590A
 ; FILING DATE: 25-JUN-1993
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Mistrock, S. Leslie
 ; REGISTRATION NUMBER: 18,872
 ; REFERENCE/DOCKET NUMBER: 7326-015
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 212 790-9090
 ; TELEFAX: 212 8698864/9741
 ; TELEX: 66141 PENNIE
 ; INFORMATION FOR SEQ ID NO: 19:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 2471 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: unknown
 ; MOLECULE TYPE: peptide
 ; US-08-083-590A-19
 ; Query Match 60.4%; Score 84; DB 1; Length 2471;
 ; Best Local Similarity 45.2%; Pred. No. 0.003;
 ; Matches 14; Conservative 2; Mismatches 15; Indels 0; Gaps 0;
 QY 1 CDXXXXXXGNGXCDXXCNNAACXXDXGDC 31
 Db 1425 CLSQYCADKARDGVCDACNSHACQMDGDC 1455
 RESULT 11
 US-08-532-384-19
 ; Sequence 19, Application US/08532384
 ; Patent No. 6083904
 ; GENERAL INFORMATION:
 ; APPLICANT: Artavanis-Tsakonas, S. et al.
 ; TITLE OF INVENTION: Therapeutic And Diagnostic Methods
 ; TITLE OF INVENTION: And Compositions Based On No. 6083904ch Proteins And
 ; TITLE OF INVENTION: Nucleic Acids
 ; NUMBER OF SEQUENCES: 21
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Pennie & Edmonds
 ; STREET: 1155 Avenue of the Americas
 ; CITY: New York
 ; STATE: New York
 ; COUNTRY: U.S.A.

ZIP: 10036
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/532,384
 ; FILING DATE:
 ; CLASSIFICATION: 424
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/083,590
 ; FILING DATE: 25-JUN-1993
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Mistrock, S. Leslie
 ; REGISTRATION NUMBER: 18,872
 ; REFERENCE/DOCKET NUMBER: 7326-015
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 212 790-9090
 ; TELEFAX: 212 8698864/9741
 ; TELEX: 66141 PENNIE
 ; INFORMATION FOR SEQ ID NO: 19:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 2471 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: unknown
 ; MOLECULE TYPE: peptide
 ; US-08-532-384-19

Query Match 60.4%; Score 84; DB 3; Length 2471;
 Best Local Similarity 45.2%; Pred. No. 0.003;
 Matches 14; Conservative 2; Mismatches 15; Indels 0; Gaps 0;
 QY 1 CDXXXXXXGNGXCDXXCNNAACXXDXGDC 31
 Db 1425 CLSQYCADKARDGVCDACNSHACQMDGDC 1455

RESULT 12
 US-08-899-232-1
 ; Sequence 1, Application US/08899232
 ; Patent No. 6436650
 ; GENERAL INFORMATION:
 ; APPLICANT: Artavanis-Tsakonas, Spyridon
 ; TITLE OF INVENTION: ACTIVATED FORMS OF NOTCH AND METHODS BASED THEREON
 ; TITLE OF INVENTION: ACTIVATED FORMS OF NOTCH AND METHODS BASED THEREON
 ; FILING DATE: 7326-046
 ; CURRENT FILING DATE: 1997-07-23
 ; SOFTWARE: Patentin Ver. 2.0
 ; SEQ ID NO: 1
 ; LENGTH: 2471
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; US-08-899-232-1

Query Match 60.4%; Score 84; DB 4; Length 2471;
 Best Local Similarity 45.2%; Pred. No. 0.003;
 Matches 14; Conservative 2; Mismatches 15; Indels 0; Gaps 0;

QY 1 CDXXXXXXGNGXCDXXCNNAACXXDXGDC 31
 Db 1425 CLSQYCADKARDGVCDACNSHACQMDGDC 1455
 RESULT 13
 US-08-185-432-17
 ; Sequence 17, Application US/08185432
 ; Patent No. 5750652
 ; GENERAL INFORMATION:
 ; APPLICANT: Artavanis-Tsakonas, Spyridon

APPLICANT: Buseau, Isabelle
APPLICANT: Diederich, Robert J.
APPLICANT: Xu, Tian
APPLICANT: Matsuno, Kenji
TITLE OF INVENTION: DELTEX PROTEINS, NUCLEIC ACIDS, AND
TITLE OF INVENTION: ANTIBODIES, AND RELATED METHODS AND COMPOSITIONS
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: PENNIE & EDMONDS
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/185,432
FILING DATE: 21-JAN-1994
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Mistrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 7326-006
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 2556 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-185-432-17

Query Match 60.4%; Score 84; DB 1; Length 2556;
Best Local Similarity 45.2%; Pred. No. 0.0031;
Matches 14; Conservative 1; Mismatches 16; Indels 0; Gaps 0;

QY 1 CDXXXKXKXGKXCDXXCNNAACXDXDC 31
DB 1450 CELPEQEDAGNKVCSLQCNHACGWDGDC 1480

RESULT 14
US-08-083-590A-20
Sequence 20, Application US/08083590A
Patent No. 5786158
GENERAL INFORMATION:
APPLICANT: Artavanis-Tsakonas, S. et al.
TITLE OF INVENTION: Therapeutic And Diagnostic Methods
TITLE OF INVENTION: And Compositions Based On No. 5786158ch Proteins And
TITLE OF INVENTION: Nucleic Acids
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/083,590A
FILING DATE: 25-JUN-1993

CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Mistrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 7326-015
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212 790-9090
TELEFAX: 212 8698864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 2556 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-083-590A-20

Query Match 60.4%; Score 84; DB 1; Length 2556;
Best Local Similarity 45.2%; Pred. No. 0.0031;
Matches 14; Conservative 1; Mismatches 16; Indels 0; Gaps 0;

QY 1 CDXXXKXKXGKXCDXXCNNAACXDXDC 31
DB 1450 CELPEQEDAGNKVCSLQCNHACGWDGDC 1480

RESULT 15
US-08-532-384-20
Sequence 20, Application US/08532384
Patent No. 6083904
GENERAL INFORMATION:
APPLICANT: Artavanis-Tsakonas, S. et al.
TITLE OF INVENTION: Therapeutic And Diagnostic Methods
TITLE OF INVENTION: And Compositions Based On No. 6083904ch Proteins And
TITLE OF INVENTION: Nucleic Acids
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/532,384
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/083,590
FILING DATE: 25-JUN-1993
ATTORNEY/AGENT INFORMATION:
NAME: Mistrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 7326-015
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212 790-9090
TELEFAX: 212 8698864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 2556 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-532-384-20

Thu May 20 14:39:13 2004

us-09-903-199-4.rai

Page 7

Query Match 60.4%; Score 84; DB 3; Length 2556;
Best Local Similarity 45.2%; Pred. No. 0.0031;
Matches 14; Conservative 1; Mismatches 16; Indels 0; Gaps 0;

```

QY      1 CDXXXCKKKKNGXCDXXCNNAACXXDXDC 31
          | | | | | | | | | | | | | |
DB      1450 CETPECQEDAGNKVCSLQCNNHACGWDGDC 1480

```

Search completed: May 19, 2004, 16:01:21
Job time : 1.29658 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 19, 2004, 15:59:55 ; Search time 3.06464 Seconds
(without alignments)
2814.722 Million cell updates/sec

Title: US-09-903-199-4

Perfect score: 139
Sequence: 1 CDXXXKXXKXNGKCDXXCNNAACXNDGDC 31

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1145568 seqs, 278261457 residues

Total number of hits satisfying chosen parameters: 1145568

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications AA:*
1: /cgn2_6/ptodata/2/pubppaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/2/pubppaa/PTCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubppaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubppaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubppaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubppaa/PTCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubppaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/2/pubppaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubppaa/US09_PUBCOMB.pep:*
10: /cgn2_6/ptodata/2/pubppaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubppaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/2/pubppaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/2/pubppaa/US10_PUBCOMB.pep:*
14: /cgn2_6/ptodata/2/pubppaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/2/pubppaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/2/pubppaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/2/pubppaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/2/pubppaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	115	82.7	31	9	US-09-903-248-4
2	115	82.7	31	9	US-09-859-604-4
3	115	82.7	31	9	US-09-903-063-4
4	115	82.7	31	9	US-09-903-216-4
5	115	82.7	31	9	US-09-903-199-4
6	115	82.7	31	9	US-09-903-023-4
7	115	82.7	31	10	US-09-436-184-4
8	105	75.5	1139	12	US-10-419-026-4
9	100	71.9	39	12	US-10-072-012-825
10	100	71.9	39	15	US-10-190-115-83
11	100	71.9	39	15	US-10-190-115-84
12	100	71.9	39	15	US-10-369-072-83
13	100	71.9	39	15	US-10-369-072-84
14	92	66.2	1429	15	US-10-369-493-5819
15	85	61.2	2531	12	US-10-072-012-470

16	85	61.2	2531	12	US-10-072-012-471	Sequence 471, App
17	85	61.2	2531	15	US-10-190-115-29	Sequence 29, Appl
18	85	61.2	2531	15	US-10-369-072-29	Sequence 29, Appl
19	84	60.4	874	15	US-10-294-006-34	Sequence 34, Appl
20	84	60.4	1015	12	US-10-419-026-1	Sequence 1, Appl1
21	84	60.4	1068	12	US-10-419-026-2	Sequence 2, Appl1
22	84	60.4	2444	9	US-09-944-849-2	Sequence 469, App
23	84	60.4	2444	12	US-10-072-012-469	Sequence 2, Appl1
24	84	60.4	2469	15	US-10-190-115-2	Sequence 2, Appl1
25	84	60.4	2469	15	US-10-369-072-2	Sequence 2, Appl1
26	84	60.4	2471	15	US-10-190-115-27	Sequence 2, Appl1
27	84	60.4	2471	15	US-10-369-072-27	Sequence 27, Appl
28	84	60.4	2555	12	US-10-072-012-468	Sequence 468, App
29	84	60.4	2556	12	US-10-072-012-134	Sequence 134, App
30	84	60.4	2556	12	US-10-072-012-467	Sequence 467, App
31	84	60.4	2556	15	US-10-294-006-12	Sequence 12, Appl
32	83	59.7	2321	14	US-10-356-625-2	Sequence 2, Appl1
33	82	59.0	2317	15	US-10-190-115-26	Sequence 26, Appl
34	82	59.0	2317	15	US-10-369-072-26	Sequence 26, Appl
35	80	57.6	1064	12	US-10-419-026-3	Sequence 3, Appl1
36	80	57.6	2524	15	US-10-190-115-25	Sequence 25, Appl
37	80	57.6	2524	15	US-10-369-072-25	Sequence 25, Appl
38	79	56.8	1295	15	US-10-369-493-5820	Sequence 5820, Ap
39	79	56.8	1295	15	US-10-369-493-5821	Sequence 5821, Ap
40	78	56.1	846	15	US-10-094-749-2332	Sequence 2392, Ap
41	78	56.1	847	15	US-10-120-801-88	Sequence 88, Appl
42	78	56.1	928	9	US-09-895-072-1	Sequence 1, Appl1
43	78	56.1	928	9	US-09-986-552-1	Sequence 1, Appl1
44	78	56.1	928	14	US-10-023-888-4	Sequence 4, Appl1
45	78	56.1	928	14	US-10-023-889-4	Sequence 4, Appl1

ALIGNMENTS

```

RESULT 1
US-09-903-248-4
? Sequence 4, Application US/09903248
? Patent No. US20020102263A1
? GENERAL INFORMATION:
? APPLICANT: Mandis, Jack R.
? APPLICANT: de la Monte, Suzanne M.
? APPLICANT: Ince, Nedim
? TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF MALIGNANT NEOPLASMS
? FILE REFERENCE: 21466-032 DIVS
? CURRENT APPLICATION NUMBER: US/09/903,248
? CURRENT FILING DATE: 2001-07-11
? PRIOR APPLICATION NUMBER: 09/436,184
? PRIOR FILING DATE: 1999-11-08
? NUMBER OF SEQ ID NOS: 9
? SOFTWARE: PatentIn Ver. 2.1
? SEQ ID NO 4
? LENGTH: 31
? TYPE: PRT
? ORGANISM: Artificial Sequence
? FEATURE:
? OTHER INFORMATION: Description of Artificial Sequence: EGF-like
? OTHER INFORMATION: cysteine-rich repeat
? NAME/KEY: VARIANT
? LOCATION: (3)..(5)
? OTHER INFORMATION: Wherein any Xaa may be any amino acid
? NAME/KEY: VARIANT
? LOCATION: (6)..(7)
? OTHER INFORMATION: Wherein Xaa is any amino acid.
? NAME/KEY: VARIANT
? LOCATION: (10)
? OTHER INFORMATION: Wherein Xaa is any amino acid.
? NAME/KEY: VARIANT
? LOCATION: (14)
? OTHER INFORMATION: Wherein Xaa is any amino acid.
? NAME/KEY: VARIANT
? LOCATION: (17)..(18)

```

```
/ NAME/KEY: VARIANT
/ LOCATION: (25)..(26)
/ OTHER INFORMATION: Wherein Xaa is any amino acid.
/ NAME/KEY: VARIANT
/ LOCATION: (29)
/ OTHER INFORMATION: Wherein Xaa is any amino acid.
US-09-903-248-4
```

```
Query Match      82.7%; Score 115; DB 9; Length 31;
Best Local Similarity 100.0%; Pred. No. 5.4e-10;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 CDXXXCXKXGKGCDDXXCNNAACXXDXDC 31
DB 1 CDXXXCXKXGKGCDDXXCNNAACXXDXDC 31
```

```
RESULT 2
US-09-859-604-4
/ Sequence 4, Application US/09859604
/ Patent No. US20020110559A1
/ GENERAL INFORMATION:
/ APPLICANT: Wands, Jack R.
/ APPLICANT: de la Monte, Suzanne M
/ APPLICANT: Deutch, Alan H
/ APPLICANT: Ghanbari, Hossein A
/ TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF MALIGNANT NEOPLASMS
/ FILE REFERENCE: 21486-032 CIP
/ CURRENT APPLICATION NUMBER: US/09/859,604
/ PRIOR FILING DATE: 2001-05-17
/ PRIOR APPLICATION NUMBER: 09/436,184
/ NUMBER OF SEQ ID NOS: 13
/ SOFTWARE: PatentIn Ver. 2.1
/ SEQ ID NO 4
/ LENGTH: 31
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Description of Artificial Sequence: EGF-like
/ NAME/KEY: VARIANT
/ LOCATION: (2)..(5)
/ OTHER INFORMATION: Wherein any Xaa may be any amino acid
/ NAME/KEY: VARIANT
/ LOCATION: (7)..(8)
/ OTHER INFORMATION: Wherein Xaa is any amino acid.
/ NAME/KEY: VARIANT
/ LOCATION: (10)
/ OTHER INFORMATION: Wherein Xaa is any amino acid.
/ NAME/KEY: VARIANT
/ LOCATION: (14)
/ OTHER INFORMATION: Wherein Xaa is any amino acid.
/ NAME/KEY: VARIANT
/ LOCATION: (17)..(18)
/ OTHER INFORMATION: Wherein Xaa is any amino acid.
/ NAME/KEY: VARIANT
/ LOCATION: (25)..(26)
/ OTHER INFORMATION: Wherein Xaa is any amino acid.
/ NAME/KEY: VARIANT
/ LOCATION: (29)
/ OTHER INFORMATION: Wherein Xaa is any amino acid.
US-09-859-604-4
```

```
Query Match      82.7%; Score 115; DB 9; Length 31;
Best Local Similarity 100.0%; Pred. No. 5.4e-10;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 CDXXXCXKXGKGCDDXXCNNAACXXDXDC 31
DB 1 CDXXXCXKXGKGCDDXXCNNAACXXDXDC 31
```

```
RESULT 3
US-09-903-063-4
```

```
/ Sequence 4, Application US/09903063
/ Patent No. US20020114810A1
/ GENERAL INFORMATION:
/ APPLICANT: Wands, Jack R.
/ APPLICANT: de la Monte, Suzanne M.
/ APPLICANT: Ince, Nedim
/ APPLICANT: Carlsson, Rolf I.
/ TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF MALIGNANT NEOPLASMS
/ FILE REFERENCE: 21486-032 DIV3
/ CURRENT APPLICATION NUMBER: US/09/903,063
/ PRIOR FILING DATE: 2001-10-11
/ PRIOR APPLICATION NUMBER: 09/436,184
/ NUMBER OF SEQ ID NOS: 9
/ SOFTWARE: PatentIn Ver. 2.1
/ SEQ ID NO 4
/ LENGTH: 31
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Description of Artificial Sequence: EGF-like
/ NAME/KEY: VARIANT
/ LOCATION: (3)..(5)
/ OTHER INFORMATION: Wherein any Xaa may be any amino acid
/ NAME/KEY: VARIANT
/ LOCATION: (6)..(7)
/ OTHER INFORMATION: Wherein Xaa is any amino acid.
/ NAME/KEY: VARIANT
/ LOCATION: (10)
/ OTHER INFORMATION: Wherein Xaa is any amino acid.
/ NAME/KEY: VARIANT
/ LOCATION: (14)
/ OTHER INFORMATION: Wherein Xaa is any amino acid.
/ NAME/KEY: VARIANT
/ LOCATION: (17)..(18)
/ OTHER INFORMATION: Wherein Xaa is any amino acid.
/ NAME/KEY: VARIANT
/ LOCATION: (25)..(26)
/ OTHER INFORMATION: Wherein Xaa is any amino acid.
/ NAME/KEY: VARIANT
/ LOCATION: (29)
/ OTHER INFORMATION: Wherein Xaa is any amino acid.
US-09-903-063-4
```

```
Query Match      82.7%; Score 115; DB 9; Length 31;
Best Local Similarity 100.0%; Pred. No. 5.4e-10;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 CDXXXCXKXGKGCDDXXCNNAACXXDXDC 31
DB 1 CDXXXCXKXGKGCDDXXCNNAACXXDXDC 31
```

```
RESULT 4
US-09-903-216-4
/ Sequence 4, Application US/09903216
/ Patent No. US20020114811A1
/ GENERAL INFORMATION:
/ APPLICANT: Wands, Jack R.
/ APPLICANT: de la Monte, Suzanne M.
/ APPLICANT: Ince, Nedim
/ APPLICANT: Carlsson, Rolf I.
/ TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF MALIGNANT NEOPLASMS
/ FILE REFERENCE: 21486-032 DIV2
/ CURRENT APPLICATION NUMBER: US/09/903,216
/ PRIOR FILING DATE: 2001-07-11
/ PRIOR APPLICATION NUMBER: 09/436,184
/ NUMBER OF SEQ ID NOS: 9
/ SOFTWARE: PatentIn Ver. 2.1
/ SEQ ID NO 4
```

```
LENGTH: 31
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: EGF-like
NAME/KEY: VARIANT
LOCATION: (3)..(5)
OTHER INFORMATION: Wherein Xaa may be any amino acid
NAME/KEY: VARIANT
LOCATION: (6)..(7)
OTHER INFORMATION: Wherein Xaa is any amino acid.
NAME/KEY: VARIANT
LOCATION: (10)
OTHER INFORMATION: Wherein Xaa is any amino acid.
NAME/KEY: VARIANT
LOCATION: (14)
OTHER INFORMATION: Wherein Xaa is any amino acid.
NAME/KEY: VARIANT
LOCATION: (17)..(18)
NAME/KEY: VARIANT
LOCATION: (25)..(26)
OTHER INFORMATION: Wherein Xaa is any amino acid.
NAME/KEY: VARIANT
LOCATION: (29)
OTHER INFORMATION: Wherein Xaa is any amino acid.
US-09-903-216-4
```

```
Query Match      82.7%; Score 115; DB 9; Length 31;
Best Local Similarity 100.0%; Pred. No. 5,4e-10;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 1 CDXXXCXKXGNGCXCDXXCNNAACXXDXGDC 31
Db 1 CDXXXCXKXGNGCXCDXXCNNAACXXDXGDC 31
```

```
RESULT 5
US-09-903-199-4
Sequence 4, Application US/09903199
Patent No. US20020122802A1
GENERAL INFORMATION:
APPLICANT: Wands, Jack R.
APPLICANT: de la Monte, Suzanne M.
APPLICANT: Ince, Nedim
APPLICANT: Carlson, Rolf I.
TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF MALIGNANT NEOPLASMS
FILE REFERENCE: 21486-032 DIV4
CURRENT APPLICATION NUMBER: US/09/903,199
CURRENT FILING DATE: 2001-07-11
PRIOR APPLICATION NUMBER: 09/436,184
NUMBER OF SEQ ID NOS: 9
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 4
LENGTH: 31
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: EGF-like
OTHER INFORMATION: cysteine-rich repeat
NAME/KEY: VARIANT
LOCATION: (3)..(5)
OTHER INFORMATION: Wherein any Xaa may be any amino acid
NAME/KEY: VARIANT
LOCATION: (6)..(7)
OTHER INFORMATION: Wherein Xaa is any amino acid.
NAME/KEY: VARIANT
LOCATION: (10)
OTHER INFORMATION: Wherein Xaa is any amino acid.
NAME/KEY: VARIANT
LOCATION: (14)
OTHER INFORMATION: Wherein Xaa is any amino acid.
NAME/KEY: VARIANT
LOCATION: (17)..(18)
OTHER INFORMATION: Wherein Xaa is any amino acid.
NAME/KEY: VARIANT
LOCATION: (25)..(26)
OTHER INFORMATION: Wherein Xaa is any amino acid.
NAME/KEY: VARIANT
LOCATION: (29)
OTHER INFORMATION: Wherein Xaa is any amino acid.
US-09-903-199-4
```

```
NAME/KEY: VARIANT
LOCATION: (17)..(18)
NAME/KEY: VARIANT
LOCATION: (25)..(26)
OTHER INFORMATION: Wherein Xaa is any amino acid.
NAME/KEY: VARIANT
LOCATION: (29)
OTHER INFORMATION: Wherein Xaa is any amino acid.
US-09-903-199-4
```

```
Query Match      82.7%; Score 115; DB 9; Length 31;
Best Local Similarity 100.0%; Pred. No. 5,4e-10;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 1 CDXXXCXKXGNGCXCDXXCNNAACXXDXGDC 31
Db 1 CDXXXCXKXGNGCXCDXXCNNAACXXDXGDC 31
```

```
RESULT 6
US-09-903-023-4
Sequence 4, Application US/09903023
Patent No. US20020146421A1
GENERAL INFORMATION:
APPLICANT: Wands, Jack R.
APPLICANT: de la Monte, Suzanne M.
APPLICANT: Ince, Nedim
APPLICANT: Carlson, Rolf I.
TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF MALIGNANT NEOPLASMS
FILE REFERENCE: 21486-032 DIV1
CURRENT APPLICATION NUMBER: US/09/903,023
CURRENT FILING DATE: 2001-10-11
PRIOR APPLICATION NUMBER: 09/436,184
NUMBER OF SEQ ID NOS: 9
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 4
LENGTH: 31
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: EGF-like
OTHER INFORMATION: cysteine-rich repeat
NAME/KEY: VARIANT
LOCATION: (3)..(5)
OTHER INFORMATION: Wherein any Xaa may be any amino acid
NAME/KEY: VARIANT
LOCATION: (6)..(7)
OTHER INFORMATION: Wherein Xaa is any amino acid.
NAME/KEY: VARIANT
LOCATION: (10)
OTHER INFORMATION: Wherein Xaa is any amino acid.
NAME/KEY: VARIANT
LOCATION: (14)
OTHER INFORMATION: Wherein Xaa is any amino acid.
NAME/KEY: VARIANT
LOCATION: (17)..(18)
OTHER INFORMATION: Wherein Xaa is any amino acid.
NAME/KEY: VARIANT
LOCATION: (25)..(26)
OTHER INFORMATION: Wherein Xaa is any amino acid.
NAME/KEY: VARIANT
LOCATION: (29)
OTHER INFORMATION: Wherein Xaa is any amino acid.
US-09-903-023-4
```

```
Query Match      82.7%; Score 115; DB 9; Length 31;
Best Local Similarity 100.0%; Pred. No. 5,4e-10;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 1 CDXXXCXKXGNGCXCDXXCNNAACXXDXGDC 31
Db 1 CDXXXCXKXGNGCXCDXXCNNAACXXDXGDC 31
```


RESULT 10
US-10-190-115-83
; Sequence 83, Application US/10190115
; Publication No. US20030207394A1

```
/ GENERAL INFORMATION:
/ APPLICANT: Alsobrook, John P. II
/ APPLICANT: Boldog, Ferenc L.
/ APPLICANT: Burgess, Catherine E.
/ APPLICANT: Casman, Stacie J.
/ APPLICANT: Grosse, William M.
/ APPLICANT: Gusev, Vladimir Y.
/ APPLICANT: Ji, Weizhen
/ APPLICANT: Lepley, Denise M.
/ APPLICANT: Liu, Xiaohong
/ APPLICANT: Mezick, Amanda J.
/ APPLICANT: Padigaru, Muralidhara
/ APPLICANT: Patturajan, Meera
/ APPLICANT: Rastelli, Luca
/ APPLICANT: Shen, Lei
/ APPLICANT: Shenoy, Suresh G.
/ APPLICANT: Shinkets, Richard A.
/ APPLICANT: Spaderna, Steven K.
/ APPLICANT: Spytek, Kimberly A.
/ APPLICANT: Szekeres, Edward S. Jr.
/ APPLICANT: Taupier, Raymond J. Jr.
/ APPLICANT: Tchernev, Velizar T.
/ APPLICANT: Zerhusen, Bryan D.
/ APPLICANT: Voss, Edward Z.
/ TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
/ FILE REFERENCE: 21402-050 CIP
/ CURRENT APPLICATION NUMBER: US/10/190,115
/ PRIOR FILING DATE: 2003-02-10
/ PRIOR APPLICATION NUMBER: 60/303,168
/ PRIOR FILING DATE: 2001-07-05
/ PRIOR APPLICATION NUMBER: 60/368,996
/ PRIOR FILING DATE: 2002-04-01
/ PRIOR APPLICATION NUMBER: 60/386,816
/ PRIOR FILING DATE: 2002-06-07
/ PRIOR APPLICATION NUMBER: 60/215,854
/ PRIOR FILING DATE: 2000-07-03
/ PRIOR APPLICATION NUMBER: 60/215,856
/ PRIOR FILING DATE: 2000-07-03
/ PRIOR APPLICATION NUMBER: 60/215,902
/ PRIOR FILING DATE: 2000-07-03
/ PRIOR APPLICATION NUMBER: 60/216,585,
/ PRIOR FILING DATE: 2000-07-07
/ PRIOR APPLICATION NUMBER: 60/216,586
/ PRIOR FILING DATE: 2001-07-07
/ PRIOR APPLICATION NUMBER: 60/216,722
/ PRIOR FILING DATE: 2000-07-07
/ PRIOR APPLICATION NUMBER: 60/218,622
/ Remaining Prior Application data removed - See File Wrapper or PALM.
/ NUMBER OF SEQ ID NOS: 136
/ SOFTWARE: Curoseqdist version 0.1
/ SEQ ID NO 83
/ LENGTH: 39
/ TYPE: PRT
/ ORGANISM: Homo sapiens
US-10-190-115-83

Query Match          71.9%; Score 100; DB 15; Length 39;
Best Local Similarity 51.6%; Pred. No. 1e-07;
Matches 16; Conservative 2; Mismatches 13; Indels 0; Gaps 0;

QY 1 CDDXXCKXKXGKXCDXXCUNNAACXXDXDC 31
Db 8 CEDAQCDKDFGDGVCDECCNNAECTLMDGDC 38

RESULT 11
US-10-190-115-84
/ Sequence 84, Application US/10190115
/ Publication No. US20030207394A1
/ GENERAL INFORMATION:
/ APPLICANT: Alsobrook, John P. II
/ APPLICANT: Boldog, Ferenc L.
```

```
/ APPLICANT: Burgess, Catherine E.
/ APPLICANT: Casman, Stacie J.
/ APPLICANT: Grosse, William M.
/ APPLICANT: Gusev, Vladimir Y.
/ APPLICANT: Ji, Weizhen
/ APPLICANT: Lepley, Denise M.
/ APPLICANT: Liu, Xiaohong
/ APPLICANT: Mezick, Amanda J.
/ APPLICANT: Padigaru, Muralidhara
/ APPLICANT: Patturajan, Meera
/ APPLICANT: Rastelli, Luca
/ APPLICANT: Shen, Lei
/ APPLICANT: Shenoy, Suresh G.
/ APPLICANT: Shinkets, Richard A.
/ APPLICANT: Spaderna, Steven K.
/ APPLICANT: Spytek, Kimberly A.
/ APPLICANT: Szekeres, Edward S. Jr.
/ APPLICANT: Taupier, Raymond J. Jr.
/ APPLICANT: Tchernev, Velizar T.
/ APPLICANT: Zerhusen, Bryan D.
/ APPLICANT: Voss, Edward Z.
/ TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
/ FILE REFERENCE: 21402-050 CIP
/ CURRENT APPLICATION NUMBER: US/10/190,115
/ PRIOR FILING DATE: 2003-02-10
/ PRIOR APPLICATION NUMBER: 60/303,168
/ PRIOR FILING DATE: 2001-07-05
/ PRIOR APPLICATION NUMBER: 60/368,996
/ PRIOR FILING DATE: 2002-04-01
/ PRIOR APPLICATION NUMBER: 60/386,816
/ PRIOR FILING DATE: 2002-06-07
/ PRIOR APPLICATION NUMBER: 60/215,854
/ PRIOR FILING DATE: 2000-07-03
/ PRIOR APPLICATION NUMBER: 60/215,856
/ PRIOR FILING DATE: 2000-07-03
/ PRIOR APPLICATION NUMBER: 60/215,902
/ PRIOR FILING DATE: 2000-07-03
/ PRIOR APPLICATION NUMBER: 60/216,585,
/ PRIOR FILING DATE: 2000-07-07
/ PRIOR APPLICATION NUMBER: 60/216,586
/ PRIOR FILING DATE: 2001-07-07
/ PRIOR APPLICATION NUMBER: 60/216,722
/ PRIOR FILING DATE: 2000-07-07
/ PRIOR APPLICATION NUMBER: 60/218,622
/ Remaining Prior Application data removed - See File Wrapper or PALM.
/ NUMBER OF SEQ ID NOS: 136
/ SOFTWARE: Curoseqdist version 0.1
/ SEQ ID NO 84
/ LENGTH: 39
/ TYPE: PRT
/ ORGANISM: Homo sapiens
US-10-190-115-84

Query Match          71.9%; Score 100; DB 15; Length 39;
Best Local Similarity 51.6%; Pred. No. 1e-07;
Matches 16; Conservative 2; Mismatches 13; Indels 0; Gaps 0;

QY 1 CDDXXCKXKXGKXCDXXCUNNAACXXDXDC 31
Db 8 CEDAQCDKDFGDGVCDECCNNAECTLMDGDC 38

RESULT 12
US-10-369-072-83
/ Sequence 83, Application US/10369072
/ Publication No. US20040014081A1
/ GENERAL INFORMATION:
/ APPLICANT: Alsobrook II, John P
/ APPLICANT: Spaderna, Steven K
/ APPLICANT: Tchernev, Velizar
/ APPLICANT: Liu, Xiaohong
/ APPLICANT: Shenoy, Suresh
```

```

; APPLICANT: Spytek, Kimberly
; APPLICANT: Zernhusen, Bryan
; APPLICANT: Paturajan, Meera
; APPLICANT: Taupier, Raymond T
; APPLICANT: Rastelli, Luca
; APPLICANT: Grosse, William M
; APPLICANT: Szereres, Edward S
; APPLICANT: Lepley, Denise M
; APPLICANT: Shen, Lei
; APPLICANT: Burgess, Catherine E
; APPLICANT: Shimeles, Richard
; APPLICANT: Padigaru, Muralidhara
; TITLE OF INVENTION: No. US20040014081A1el Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-050 CON2
; CURRENT APPLICATION NUMBER: US/10/369,072
; PRIOR FILING DATE: 2003-02-18
; PRIOR APPLICATION NUMBER: 10/174,372
; PRIOR FILING DATE: 2002-06-17
; PRIOR APPLICATION NUMBER: 09/898,994
; PRIOR FILING DATE: 2001-07-03
; PRIOR APPLICATION NUMBER: 60/215,854
; PRIOR FILING DATE: 2000-07-03
; PRIOR APPLICATION NUMBER: 60/215,856
; PRIOR FILING DATE: 2000-07-03
; PRIOR APPLICATION NUMBER: 60/215,902
; PRIOR FILING DATE: 2000-07-03
; PRIOR APPLICATION NUMBER: 60/216,585
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: 60/216,586
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: 60/216,722
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: 60/218,622
; PRIOR FILING DATE: 2000-07-17
; PRIOR APPLICATION NUMBER: 60/218,992
; PRIOR FILING DATE: 2000-07-17
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 100
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 83
; LENGTH: 39
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Domain search
; US-10-369-072-83

Query Match
Best Local Similarity 71.9%; Score 100; DB 15; Length 39;
Matches 16; Conservative 2; Mismatches 13; Indels 0; Gaps 0;

QY 1 CDXXXCXXKXGNGXCDXXCNNAACXXDXDC 31
DB 8 CEDAQCMDFGdGVCDCECNNAECIMDGDGC 38

RESULT 13
US-10-369-072-84
; Sequence 84, Application US/10369072
; Publication No. US20040014081A1
; GENERAL INFORMATION:
; APPLICANT: Alsobrook II, John P
; APPLICANT: Spaderma, Stephen K
; APPLICANT: Tchernev, Velizar
; APPLICANT: Liu, Xiaohong
; APPLICANT: Shenoy, Suresh
; APPLICANT: Spytek, Kimberly
; APPLICANT: Zernhusen, Bryan
; APPLICANT: Paturajan, Meera
; APPLICANT: Taupier, Raymond T
; APPLICANT: Rastelli, Luca
; APPLICANT: Grosse, William M
```

```

; APPLICANT: Szereres, Edward S
; APPLICANT: Lepley, Denise M
; APPLICANT: Shen, Lei
; APPLICANT: Burgess, Catherine B
; APPLICANT: Shimeles, Richard
; APPLICANT: Padigaru, Muralidhara
; TITLE OF INVENTION: No. US20040014081A1el Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-050 CON2
; CURRENT APPLICATION NUMBER: US/10/369,072
; PRIOR FILING DATE: 2003-02-18
; PRIOR APPLICATION NUMBER: 10/174,372
; PRIOR FILING DATE: 2002-06-17
; PRIOR APPLICATION NUMBER: 09/898,994
; PRIOR FILING DATE: 2001-07-03
; PRIOR APPLICATION NUMBER: 60/215,854
; PRIOR FILING DATE: 2000-07-03
; PRIOR APPLICATION NUMBER: 60/215,856
; PRIOR FILING DATE: 2000-07-03
; PRIOR APPLICATION NUMBER: 60/215,902
; PRIOR FILING DATE: 2000-07-03
; PRIOR APPLICATION NUMBER: 60/216,585
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: 60/216,586
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: 60/216,722
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: 60/218,622
; PRIOR FILING DATE: 2000-07-17
; PRIOR APPLICATION NUMBER: 60/218,992
; PRIOR FILING DATE: 2000-07-17
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 100
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 84
; LENGTH: 39
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Domain search
; US-10-369-072-84

Query Match
Best Local Similarity 71.9%; Score 100; DB 15; Length 39;
Matches 16; Conservative 2; Mismatches 13; Indels 0; Gaps 0;

QY 1 CDXXXCXXKXGNGXCDXXCNNAACXXDXDC 31
DB 8 CEDAQCMDFGdGVCDCECNNAECIMDGDGC 38

RESULT 14
US-10-369-493-5819
; Sequence 5819, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 5819
; LENGTH: 1429
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
```

US-10-369-493-5819

Query Match 66.2%; Score 92; DB 15; Length 1429;
Best Local Similarity 45.2%; Pred. No. 4e-05;
Matches 14; Conservative 3; Mismatches 14; Indels 0; Gaps 0;

QY 1 CDXXXCKXKXGKXGKXCDXCNNAACXDXGDC 31
DB 638 CEKRCSEBANDGNCADCVYAAACKFDGDC 668

RESULT 15
US-10-072-012-470

Sequence 470; Application US/10072012
Publication No. US2004003493A1

GENERAL INFORMATION:

APPLICANT: Tchervet, Velizar

APPLICANT: Spytek, Kimberly

APPLICANT: Zernhusen, Bryan

APPLICANT: Patunajan, Weera

APPLICANT: Shinkels, Richard

APPLICANT: Li, Li

APPLICANT: Gangoli, Esha

APPLICANT: Padigaru, Murajidhara

APPLICANT: Anderson, David W.

APPLICANT: Rastelli, Luca

APPLICANT: Miller, Charles E.

APPLICANT: Gerlach, Valerie

APPLICANT: Taupier Jr, Raymond J.

APPLICANT: Gusev, Vladimir Y.

APPLICANT: Coleman, Steven D.

APPLICANT: Wolenc, Adam R.

APPLICANT: Pena, Carol E. A

APPLICANT: Furtak, Katarzyna

APPLICANT: Grosse, William M.

APPLICANT: Alsobrook II, John P.

APPLICANT: Lepley, Denise M.

APPLICANT: Rieger, Daniel K.

APPLICANT: Burgess, Catherine E.

TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same

FILE REFERENCE: 21402-258

CURRENT APPLICATION NUMBER: US/10/072,012

PRIOR FILING DATE: 2002-01-31

PRIOR APPLICATION NUMBER: 60/265,102

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: 60/265,514

PRIOR FILING DATE: 2001-01-31

PRIOR APPLICATION NUMBER: 60/265,517

PRIOR FILING DATE: 2001-01-31

PRIOR APPLICATION NUMBER: 60/265,412

PRIOR FILING DATE: 2001-01-31

PRIOR APPLICATION NUMBER: 60/265,395

PRIOR FILING DATE: 2001-01-31

PRIOR APPLICATION NUMBER: 60/266,406

PRIOR FILING DATE: 2001-02-02

PRIOR APPLICATION NUMBER: 60/266,767

PRIOR FILING DATE: 2001-02-05

PRIOR APPLICATION NUMBER: 60/267,057

PRIOR FILING DATE: 2001-02-07

PRIOR APPLICATION NUMBER: 60/266,975

PRIOR FILING DATE: 2001-02-07

PRIOR APPLICATION NUMBER: 60/267,459

PRIOR FILING DATE: 2001-02-08

REMAINING PRIOR APPLICATION DATA REMOVED - See File Wrapper or PALM.

NUMBER OF SEQ ID NOS: 1391

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 470

LENGTH: 2531

TYPE: PRT

ORGANISM: Rattus norvegicus

US-10-072-012-470

Query Match 61.2%; Score 85; DB 12; Length 2531;

Best Local Similarity 45.2%; Pred. No. 0.00071;
Matches 14; Conservative 2; Mismatches 15; Indels 0; Gaps 0;

QY 1 CDXXXCKXKXGKXGKXCDXCNNAACXDXGDC 31
DB 1449 CELPECCQEDAGNRCVNLQCNNAACGWDGDC 1479

Search completed: May 19, 2004, 16:09:34
Job time: 3.06464 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 19, 2004, 15:54:54 ; Search time 1.29658 Seconds
(without alignments)
2299.852 Million cell updates/sec

Title: US-09-903-199-4
Perfect score: 139
Sequence: 1 CDXXXCKXKXNGKCDXXCNNAACXDXDC 31

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues
Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: PIR.78:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length DB	ID	Description
1	105	75.5	2703	1 A24420	notch protein - fr
2	97	69.8	2437	2 S42612	transmembrane prot
3	92	66.2	1429	2 S06434	homeotic protein 1
4	91	65.5	2531	2 T31070	notch homolog - se
5	88	63.3	1203	2 A49175	Notch B protein -
6	85	61.2	387	2 B49175	Notch A protein -
7	85	61.2	861	2 A48825	Notch homolog Motc
8	85	61.2	2531	2 A18188	notch-1 protein -
9	85	61.2	2531	2 A46019	cell-fate determin
10	84	60.4	2471	2 A49128	notch protein homo
11	84	60.4	2555	2 A40043	notch protein - h
12	83	59.7	2321	2 S78549	notch 3 protein -
13	82	59.0	2318	2 S45306	Notch protein - Af
14	80	57.6	2524	2 A35844	Notch protein - Af
15	79	56.8	1295	2 A32901	Notch protein - Af
16	79	56.8	1964	2 T09059	Notch protein - Af
17	77	55.4	2352	2 T30201	Notch protein - Af
18	61	43.9	67	2 T11347	Notch homolog prot
19	60	43.2	1627	2 S65464	metallothionein -
20	52	37.4	2533	2 T28675	pregnancy-associated
21	52	37.4	2533	2 T28675	alpha-51D immobili
22	50	36.0	65	2 A38739	metallothionein -
23	49	35.3	2543	2 T31687	metallothionein -
24	48.5	34.9	616	2 A55796	surface precursor - P
25	47.5	34.2	2395	2 S50820	Notch protein - P
26	46.5	33.5	419	2 A59414	metalloproteinase
27	46.5	33.5	571	2 S24789	metalloproteinase
28	46.5	33.5	609	2 S55270	cathepsin C precu
29	46	33.1	575	1 THHUB	cathepsin C precu

30	45	32.4	64	2 A33825	metallothionein A
31	45	32.4	71	2 A59412	KGD-bearing platelet
32	45	32.4	72	2 B43019	platelet aggregati
33	45	32.4	72	2 D43019	platelet aggregati
34	45	32.4	73	2 A43019	platelet aggregati
35	45	32.4	73	2 C43019	platelet aggregati
36	45	32.4	73	2 B40003	platelet aggregati
37	45	32.4	73	2 A40003	platelet aggregati
38	45	32.4	2706	2 T28155	variant-specific s
39	45	32.4	2907	2 A57278	fibrillin-2 precu
40	45	32.4	2918	2 A54105	fibrillin-2 precu
41	44.5	32.0	64	2 A25775	metallothionein A
42	44.5	32.0	313	2 S44208	extracellular matr
43	44.5	32.0	788	2 A37057	integrin beta-6 ch
44	44	31.7	186	2 A28401	agglutinin isolat
45	44	31.7	673	2 A48089	growth arrest-spec

ALIGNMENTS

RESULT 1
A24420
notch protein - fruit fly (Drosophila melanogaster)
N/Alternate names: neurogenic repetitive locus protein
C/Species: Drosophila melanogaster
C/Date: 10-Sep-1999 #sequence revision 10-Sep-1999 #text_change 10-Sep-1999
C/Accession: A24420; A24768; S09358; A05267
R/Kidd, S.; Kelley, M.R.; Young, M.W.
Mol. Cell. Biol. 6, 3094-3108, 1986
A/Reference number: A24420; MUID:87064624; PMID:3097517
A/Accession: A24420
A/Molecule type: DNA
A/Residues: 1-2703 <RID>
A/Cross-references: GB:R03508; NID:g157991; PID:AAA28725.1; PID:g157993
R/Marion, K.A.; Johansen, K.M.; Xu, T.; Aravanis-Tsakonas, S.
Cell 43, 567-581, 1985
A/Reference number: A24768; MUID:86079539; PMID:3935325
A/Accession: A24768
A/Molecule type: mRNA
A/Residues: 1-48; 'I', 50-118; 'R', 120-230; 'I', 232-256; 'N', 258-266; 'A', 268-872; 'R', 874-958
A/Note: the authors translated the codon ATC for residue 49 as Thr, ATT for residue 204
R/Butz, D.
Nucleic Acids Res. 17, 6463-6471, 1989
A/Title: Hypervariability of simple sequences as a general source for polymorphic DNA m
A/Reference number: S09358; MUID:89385974; PMID:2780284
A/Accession: S09358
A/Molecule type: DNA
A/Residues: 2505-2551; 'QQQ', 2552-2576; 'E', 2578-2604 <TAU>
R/Marion, K.A.; Yedvobnick, B.; Finnerty, V.G.; Aravanis-Tsakonas, S.
Cell 40, 55-62, 1985
A/Title: opa: a novel family of transcribed repeats shared by the Notch locus and other
A/Reference number: A05267; MUID:85093329; PMID:2981631
A/Accession: A05267
A/Molecule type: DNA
A/Residues: 2504-2576; 'E', 2578-2611 <WMA>
C/Genetics:
A/Gene: notch; opa
A/Cross-references: FlyBase:FBgn0004647
A/Map position: 8.96-9.36
A/Intons: 53/3; 84/3; 171/3; 240/3; 283/3; 2333/3; 2436/3; 2588/3
C/Superfamily: notch protein; ankyrin repeat homology; EGF homology
C/Keywords: differentiation; tandem repeat; transmembrane protein
F/27-43/Domain: transmembrane #status predicted <TMW1>
F/297-328/Domain: EGF homology <EGX1>
F/530-561/Domain: EGF homology <EGF1>
F/568-599/Domain: EGF homology <EGF>
F/988-1019/Domain: EGF homology <EGX2>
F/1064-1095/Domain: EGF homology <EGF3>
F/1187-1218/Domain: EGF homology <EGF3>
F/1746-1765/Domain: transmembrane #status predicted <TMW2>
F/1950-1989/Domain: ankyrin repeat homology <ANI>
F/1983-2015/Domain: ankyrin repeat homology <AN2>

F:1988-2004/Domain: transmembrane #status predicted <TM>
 F:2017-2049/Domain: ankyrin repeat homology <AN3>
 F:2050-2082/Domain: ankyrin repeat homology <AN4>
 F:2083-2115/Domain: ankyrin repeat homology <AN5>
 F:2538-2568/Region: glutamine-rich
 F:2538-2568/Domain: neurogenic repetitive element #status predicted <OPA>

Query Match 75.5%; Score 105; DB 1; Length 2703;
 Best Local Similarity 54.8%; Pred. No. 3.1e-07;
 Matches 17; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 1 CDXXXCXKXGNGXCDXXCNNAACXXDXDC 31
 DB 1482 CDKRGCTEKQNGICSDCNTYACNFDGDC 1512

RESULT 2
 S42612
 Transmembrane protein precursor - zebra fish

C:Species: Brachydanio rerio (zebra fish)
 C:Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 02-Aug-2002
 C:Accession: S42612
 R:Biorkamp, C.; Campos-Ortega, J.A.
 Mech. Dev. 43, 87-100, 1993

A:Title: A zebrafish homologue of the Drosophila neurogenic gene Notch and its pattern c
 A:Reference number: S42612; PMID:94128602; PMID:8297791

A:Accession: S42612
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-2437 <BIE>

A:Cross-references: EMBL:X69088; NID:9433866; PIDN:CAA48831.1; PID:9433867
 C:Superfamily: notch protein; ankyrin repeat homology; EGF homology

F:1023-1054/Domain: EGF homology <EGF1>
 F:1185-1216/Domain: EGF homology <EGF2>
 F:1915-1947/Domain: ankyrin repeat homology <AN1>
 F:1948-1980/Domain: ankyrin repeat homology <AN2>
 F:1982-2014/Domain: ankyrin repeat homology <AN3>
 F:2015-2047/Domain: ankyrin repeat homology <AN4>
 F:2048-2080/Domain: ankyrin repeat homology <AN5>

Query Match 69.8%; Score 97; DB 2; Length 2437;
 Best Local Similarity 48.4%; Pred. No. 3.9e-06;
 Matches 15; Conservative 2; Mismatches 14; Indels 0; Gaps 0;

QY 1 CDXXXCXKXGNGXCDXXCNNAACXXDXDC 31
 DB 1447 CEIAQCEGRGNALCDTQCNHACGMDGDC 1477

RESULT 3
 S06434
 homeotic protein lin-12 precursor - Caenorhabditis elegans

C:Species: Caenorhabditis elegans
 C:Date: 29-Jan-1993 #sequence_revision 29-Jan-1993 #text_change 20-Sep-1999
 C:Accession: S06434; A24769
 R:Yochem, J.; Weston, K.; Greenwald, I.
 Nature 335, 547-550, 1988

A:Title: The Caenorhabditis elegans lin-12 gene encodes a transmembrane protein with ove
 A:Reference number: S06434; PMID:88334747; PMID:3419531

A:Accession: S06434
 A:Molecule type: DNA
 A:Residues: 1-1429 <YOC>

A:Cross-references: EMBL:M12069; NID:9156357; PIDN:AAA0191.1; PID:9156356
 R:Greenwald, I.

Cell 43, 583-590, 1985
 A:Reference number: A24769; PMID:86079540; PMID:3000611

A:Accession: A24769
 A:Molecule type: DNA
 A:Residues: 173-712 <GRE>

C:Genetics: 50/2; 90/1; 109/1; 172/3; 545/1; 589/2; 632/2; 1273/3; 1389/3
 A:Mutations: 50/2; 90/1; 109/1; 172/3; 545/1; 589/2; 632/2; 1273/3; 1389/3

C:Superfamily: unassigned ankyrin repeat proteins; ankyrin repeat homology; EGF homology

C:Keywords: glycoprotein; transmembrane protein
 F:254-284/Domain: EGF homology <EGF1>
 F:507-540/Domain: EGF homology <EGF2>
 F:547-578/Domain: EGF homology <EGF3>
 F:909-931/Domain: transmembrane #status predicted <TM>
 F:1093-1125/Domain: ankyrin repeat homology <AN1>
 F:1206-1238/Domain: ankyrin repeat homology <AN2>
 F:1240-1272/Domain: ankyrin repeat homology <AN3>

Query Match 66.2%; Score 92; DB 2; Length 1429;
 Best Local Similarity 45.2%; Pred. No. 1.4e-05;
 Matches 14; Conservative 3; Mismatches 14; Indels 0; Gaps 0;

QY 1 CDXXXCXKXGNGXCDXXCNNAACXXDXDC 31
 DB 638 CEKRCSEKANDGNCADONVACCFDGDGC 668

RESULT 4
 T31070
 notch homology - sea urchin (Lytechinus variegatus)

C:Species: Lytechinus variegatus (variegated urchin)
 C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 31-Jan-2000
 C:Accession: T31070
 R:Sherwood, D.R.; McClay, D.R.
 Development 124, 3363-3374, 1997

A:Title: Identification and localization of a sea urchin Notch homologue: insights into
 A:Reference number: Z20966; PMID:97454256; PMID:9310331

A:Accession: T31070
 A:Status: preliminary; translated from GB/EMBL/DDBU
 A:Molecule type: mRNA

A:Residues: 1-2531 <SHE>
 A:Cross-references: EMBL:AF00634; NID:92570350; PID:92570351; PIDN:AA82088.1

C:Superfamily: notch protein; ankyrin repeat homology; EGF homology

Query Match 65.5%; Score 91; DB 2; Length 2511;
 Best Local Similarity 53.8%; Pred. No. 2.9e-05;
 Matches 14; Conservative 1; Mismatches 11; Indels 0; Gaps 0;

QY 6 CXXKXGNGXCDXXCNNAACXXDXDC 31
 DB 1526 CLERYGNGFCDERBCNNTGCLYDGDGC 1551

RESULT 5
 A49175
 Notch B protein - mouse (fragment)

N:Alternate names: Notch homolog
 C:Species: Mus musculus (house mouse)
 C:Date: 21-Jan-1994 #sequence_revision 05-Jan-1996 #text_change 08-Sep-2002
 C:Accession: A49175; PH1570; S32113
 R:Lardelli, M.; Lendahl, U.
 Exp. Cell Res. 204, 364-372, 1993

A:Title: Notch A and Notch B--two mouse Notch homologues coexpressed in a wide variety
 A:Reference number: A49175; PMID:33178563; PMID:8440332

A:Accession: A49175
 A:Status: preliminary; nucleic acid sequence not shown

A:Molecule type: mRNA
 A:Residues: 1-1203 <LAR>

A:Cross-references: EMBL:X68279; NID:9287989; PIDN:CAA48340.1; PID:9287990
 A:Experimental source: embryo

A:Note: sequence extracted from NCBI backbone (NCBIP:126158)

C:Comment: This protein has many EGF repeats and lin-12/Notch repeats.
 C:Comment: This protein is one of the neurogenic proteins controlling the decision between

C:Superfamily: notch protein; ankyrin repeat homology; EGF homology
 F:143-174/Domain: EGF homology <EGX1>
 F:148-513/Domain: EGF homology <EGF1>
 F:560-591/Domain: EGF homology <EGF2>
 F:674-705/Domain: EGF homology <EGX2>
 F:712-743/Domain: EGF homology <EGF3>
 F:836-867/Domain: EGF homology <EGX3>

Query Match 63.3%; Score 88; DB 2; Length 1203;

Best Local Similarity 45.2%; Pred. No. 4.5e-05;
Matches 14; Conservative 2; Mismatches 15; Indels 0; Gaps 0;

Oy 1 CDXXXXXXGNGCXCDXXCNNAACXDXGDC 31
Db 1108 CQSOYCADKARBDICDEACNSHACMGDGDGDC 1138

RESULT 6

Notch A protein - mouse (fragment)
B49175
N:Alternate names: Notch homolog
C:Species: Mus musculus (house mouse)
C:Date: 21-Jan-1994 #sequence_revision 05-Jan-1996 #text_change 08-Sep-2002
C:Accession: B49175; PH1569; S32109
R:Lardelli, M.; Lendahl, U.
Exp. Cell Res. 204, 364-372, 1993
A:Title: Notch A and Notch B--two mouse Notch homologues coexpressed in a wide variety of
A:Reference number: A49175; MUID:93178563; PMID:8440332
A:Accession: B49175
A:Status: preliminary; nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-387 <LARD>
A:Cross-references: EMBL:X68278; NID:g287987; PIDN:CAA48339.1; PID:g287988
A:Experimental source: embryo
A:Note: sequence extracted from NCBI backbone (NCBIF:126159)
C:Comment: This protein has many EGF repeats and 11-12/Notch repeats.
C:Comment: This protein is one of the neurogenic proteins controlling the decision between
F:27-58/Domain: notch protein; ankyrin repeat homology; EGF homology
F:73-104/Domain: EGF homology <EGF2>
F:151-185/Domain: EGF homology <EGF1>

Query Match 61.2%; Score 85; DB 2; Length 387;
Best Local Similarity 45.2%; Pred. No. 5.3e-05;
Matches 14; Conservative 2; Mismatches 15; Indels 0; Gaps 0;

Oy 1 CDXXXXXXGNGCXCDXXCNNAACXDXGDC 31
Db 289 CELPECCVDAGNKVCNLCQNNHACMGDGDGDC 319

RESULT 7

Notch homolog Notch protein - mouse (fragment)
A48825
C:Species: Mus musculus (house mouse)
C:Date: 01-Dec-1993 #sequence_revision 18-Nov-1994 #text_change 08-Sep-2002
C:Accession: A48825
R:Resume, A.G.; Conlon, R.A.; Ziringibl, R.; Yamaguchi, T.P.; Roseant, J.
Dev. Biol. 154, 377-387, 1992
A:Title: Expression analysis of a Notch homologue in the mouse embryo.
A:Reference number: A48825; MUID:93050801; PMID:1426644
A:Accession: A48825
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-861 <REA>
A:Experimental source: embryo
A:Note: sequence extracted from NCBI backbone (NCBIF:119144)
C:Superfamily: notch protein; ankyrin repeat homology; EGF homology
F:26-57/Domain: EGF homology <EGF>
F:64-95/Domain: EGF homology <EGX1>
F:198-229/Domain: EGF homology <EGF2>
F:441-472/Domain: EGF homology <EGX2>

Query Match 61.2%; Score 85; DB 2; Length 861;
Best Local Similarity 45.2%; Pred. No. 9.4e-05;
Matches 14; Conservative 2; Mismatches 15; Indels 0; Gaps 0;

Oy 1 CDXXXXXXGNGCXCDXXCNNAACXDXGDC 31
Db 630 CELPECCVDAGNKVCNLCQNNHACMGDGDGDC 660

RESULT 8

Notch protein homolog - rat
S18188
C:Species: Rattus norvegicus (Norway rat)
C:Date: 19-Feb-1994 #sequence_revision 10-Nov-1995 #text_change 02-Aug-2002
C:Accession: S18188
R:Weinmaster, G.; Roberts, V.J.; Lemke, G.
Development 113, 199-205, 1991
A:Title: A homolog of Drosophila Notch expressed during mammalian development.
A:Reference number: S18188; MUID:92111383; PMID:1764995
A:Accession: S18188
A:Molecule type: mRNA
A:Residues: 1-2531 <WEI>
A:Cross-references: EMBL:X57405; NID:g57634; PID:g57635
C:Superfamily: notch protein; ankyrin repeat homology; EGF homology
F:987-1018/Domain: EGF homology <EGF1>
F:1025-1056/Domain: EGF homology <EGF2>
F:1233-1264/Domain: EGF homology <EGF3>
F:1917-1949/Domain: ankyrin repeat homology <AN1>
F:1950-1982/Domain: ankyrin repeat homology <AN2>
F:1984-2016/Domain: ankyrin repeat homology <AN3>
F:2017-2049/Domain: ankyrin repeat homology <AN4>
F:2050-2082/Domain: ankyrin repeat homology <AN5>

Query Match 61.2%; Score 85; DB 2; Length 2531;
Best Local Similarity 45.2%; Pred. No. 0.0002;
Matches 14; Conservative 2; Mismatches 15; Indels 0; Gaps 0;

Oy 1 CDXXXXXXGNGCXCDXXCNNAACXDXGDC 31
Db 1449 CELPECCVDAGNKVCNLCQNNHACMGDGDGDC 1479

RESULT 9

Notch-1 protein - mouse
A46019
N:Alternate names: notch protein
C:Species: Mus musculus (house mouse)
C:Date: 22-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 07-Mar-2003
C:Accession: A46019; S25144; C49175; B46438; A46438; PH1569; S32109
R:del Amo, F.F.; Gendron-Maguire, M.; Swatek, P.J.; Jenkins, N.A.; Copeland, N.G.; Gri
Genomics 15, 259-264, 1993
A:Title: Cloning, analysis, and chromosomal localization of Notch-1, a mouse homolog of
A:Reference number: A46019; MUID:93194170; PMID:8449489
A:Accession: A46019
A:Status: not compared with conceptual translation
A:Molecule type: nucleic acid
A:Residues: 1-2531
A:Cross-references: GB:Z11886; GB:S47228; NID:g288502; PIDN:CAA77941.1; PID:g288503
A:Note: sequence extracted from NCBI backbone (NCBIF:127318)
R:Francia del Amo, F.; Smith, D.E.; Swatek, P.J.; Gendron-Maguire, M.; Greenspan, R.J.;
submitted to the EMBL Data Library, April 1992
A:Description: Expression pattern of Notch, a mouse homolog of Drosophila Notch, sugges

A:Reference number: S25144
A:Accession: S25144
A:Molecule type: mRNA
A:Residues: 1551-2108, 'O', 2110-2114, 'ALP', 2118-2170 <FRA>
A:Cross-references: EMBL:Z11886
R:Lardelli, M.; Lendahl, U.
Exp. Cell Res. 204, 364-372, 1993
A:Title: Notch A and Notch B--two mouse Notch homologues coexpressed in a wide variety
A:Reference number: A49175; MUID:93178563; PMID:8440332
A:Accession: C49175
A:Status: preliminary; nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1161-1547 <LARD>
A:Cross-references: EMBL:X68278; NID:g287987; PIDN:CAA48339.1; PID:g287988
A:Experimental source: embryo
A:Note: sequence extracted from NCBI backbone (NCBIF:126159)
R:Kopan, R.; Weintraub, H.
J. Cell Biol. 121, 631-641, 1993
A:Title: Mouse notch expression in hair follicles correlates with cell fate determinat
A:Reference number: A46438; MUID:93252998; PMID:8486742

A:Accession: R46438
A:Status: preliminary
A:Molecule type: nucleic acid
A:Residues: 1865-1932,'RR',1935-1937,'L',1938-1967,'I',1969-2044,'IE',2047-2052,'S',2053-2054,'E'
A:Experimental source: embryo
A:Note: Sequence extracted from NCBI backbone (NCBI:131246, NCBI:P:131247)
C:Comment: This protein has many EGF repeats and 11n-12[1172]/Notch repeats.
C:Comment: This protein is one of the neurogenic proteins controlling the decision between C:Genetics:
A:Gene: notch-1
A:Map position: 2
A:Note: proximal region of chromosome 2
C:Superfamily: notch protein; ankyrin repeat homology; EGF homology
F:106-118/Domain: EGF homology <EGF1>
F:114-175/Domain: EGF homology <EG01>
F:122-254/Domain: EGF homology <EGF2>
F:261-292/Domain: EGF homology <EG02>
F:339-370/Domain: EGF homology <EG03>
F:416-449/Domain: EGF homology <EGF3>
F:456-487/Domain: EGF homology <EG04>
F:494-525/Domain: EGF homology <EG05>
F:533-563/Domain: EGF homology <EG06>
F:607-638/Domain: EGF homology <EG07>
F:682-713/Domain: EGF homology <EG08>
F:757-788/Domain: EGF homology <EG09>
F:795-826/Domain: EGF homology <EG10>
F:873-904/Domain: EGF homology <EG11>
F:911-942/Domain: EGF homology <EG12>
F:949-980/Domain: EGF homology <EG13>
F:987-1018/Domain: EGF homology <EG14>
F:1025-1056/Domain: EGF homology <EG15>
F:1063-1094/Domain: EGF homology <EG16>
F:1149-1180/Domain: EGF homology <EG17>
F:1187-1218/Domain: EGF homology <EG18>
F:1233-1264/Domain: EGF homology <EGF4>
F:1352-1383/Domain: EGF homology <EG19>
F:1391-1425/Domain: EGF homology <EGF>
F:1397-1948/Domain: ankyrin repeat homology <AN1>
F:1949-1981/Domain: ankyrin repeat homology <AN2>
F:1983-2015/Domain: ankyrin repeat homology <AN3>
F:2016-2048/Domain: ankyrin repeat homology <AN4>
F:2049-2081/Domain: ankyrin repeat homology <AN5>

Query Match 61.2% Score 85; DB 2; Length 2531;
Best Local Similarity 45.2% Pred. No. 0.0002;
Matches 14; Conservative 2; Mismatches 15; Indels 0; Gaps 0;

1 CDXXXXXXKXGKXCDXXCNNAACXDXGDC 31
Db 1449 CELPEQVDAGNKVCNLCNNHACGMDGDC 1479

RESULT 10
A49128
Cell-fate determining gene Notch2 protein - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 21-Jan-1994 #sequence_revision 18-Nov-1994 #text_change 02-Aug-2002
C:Accession: A49128
R:Weinmaster, G.; Roberts, V.J.; Lemke, G.
Development 116, 931-941, 1992
A:Title: Notch2: a second mammalian Notch gene.
A:Reference number: A49128; PMID:93202015; PMID:1255745
A:Accession: A49128
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-2471 <ME1>
A:Experimental source: Schwann cell
C:Superfamily: notch protein; ankyrin repeat homology; EGF homology
F:264-295/Domain: EGF homology <EGX1>
F:799-830/Domain: EGF homology <EGF1>
F:877-908/Domain: EGF homology <EGX2>
F:1029-1060/Domain: EGF homology <EGF>

```

F:1067-1098/Domain: EGF homology <EGX3>
F:1153-1184/Domain: EGF homology <EGF3>
F:1191-1222/Domain: EGF homology <EGF4>
F:1876-1908/Domain: ankyrin repeat homology <AN1>
F:1909-1941/Domain: ankyrin repeat homology <AN2>
F:1943-1975/Domain: ankyrin repeat homology <AN3>
F:1976-2008/Domain: ankyrin repeat homology <AN4>
F:2009-2041/Domain: ankyrin repeat homology <AN5>

Query Match
Best Local Similarity 60.4%; Score 84; DB 2; Length 2471;
Matches 14; Conservative 2; Mismatches 15; Indels 0; Gaps 0;

QY
1 CDXXXCKXKXGKCDXXCNNAACXDXGDC 31
1425 CLSGYCADKARDGICDEACNSHACQMGDGD 1455

RESULT 11
A40043
notch protein homolog TAN-1 precursor - human
C:Species: Homo sapiens (man)
C:Date: 21-Apr-1992 #sequence_revision 21-Apr-1992 #text_change 02-Aug-2002
C:Accession: A40043
R:Ellisen, L.W.; Bird, J.; West, D.C.; Sorens, A.L.; Reynolds, T.C.; Smith, S.D.; Sklar
Cell 66, 649-661, 1991
A>Title: TAN-1, the human homolog of the Drosophila Notch gene, is broken by chromosome
A:Reference number: A40043; MUID:91347367; PMID:1831692
A:Accession: A40043
A>Status: preliminary; nucleic acid sequence not shown; not compared with conceptual tr
A:Molecule type: mRNA
A:Residues: 1-2555 <EHL>
A:Cross-references: GB:M73980
C:Superfamily: notch protein; ankyrin repeat homology; EGF homology
F:261-292/Domain: EGF homology <EGX1>
F:494-525/Domain: EGF homology <EGF1>
F:987-1018/Domain: EGF homology <EGX2>
F:1149-1180/Domain: EGF homology <EGF>
F:1187-1264/Domain: EGF homology <EGF3>
F:1233-1264/Domain: EGF homology <EGX3>
F:1927-1959/Domain: ankyrin repeat homology <AN1>
F:1960-1992/Domain: ankyrin repeat homology <AN2>
F:1994-2026/Domain: ankyrin repeat homology <AN3>
F:2027-2059/Domain: ankyrin repeat homology <AN4>
F:2060-2092/Domain: ankyrin repeat homology <AN5>

Query Match
Best Local Similarity 60.4%; Score 84; DB 2; Length 2555;
Matches 14; Conservative 1; Mismatches 16; Indels 0; Gaps 0;

QY
1 CDXXXCKXKXGKCDXXCNNAACXDXGDC 31
Db
1449 CELPECEGDAGNKKVCSLCCNNHACGWDGDC 1479

RESULT 12
S78549
notch3 protein - human
C:Species: Homo sapiens (man)
C:Date: 24-Jul-1998 #sequence_revision 24-Jul-1998 #text_change 08-Sep-2002
C:Accession: S78549; S71825
R:Joutel, A.; Corpechot, C.; Ducros, A.; Vahedi, K.; Chabriat, H.; Mouton, P.; Alamowit
X: M.M.; Weissenbach, J.; Bach, J.F.; Boussier, M.G.; Tournier-Lasserre, E.
Nature 383, 707-710, 1996
A:Reference number: S71825; MUID:97032728; PMID:8878478
A:Accession: S78549
A:Molecule type: mRNA
A:Residues: 1-2321 <OOU>
A:Cross-references: EMBL:U97669; NID:g2668591; PIDN:AB91371.1; PID:g2668592

```


A:Accession: S71825
 A:Status: nucleic acid sequence not shown
 A:Molecule type: DNA
 A:Residues: 67-113;138-194;268-333; 'G', 335-346;536-613;716-765;1240-1279;1815-1888 <J0012
 A/Cross-references: EMBL:U97669
 C/Genetics:
 A:Gene: notch3
 A:Map position: 19p13.1
 C/Function:
 A:Description: may be involved in pathogenesis of CADASIL, causing a type of stroke and
 C/Superfamily: notch protein; ankyrin repeat homology; EGF homology
 C/Keywords: tandem repeat; transmembrane protein
 F:123-155/Domain: EGF homology <EGF1>
 F:162-194/Domain: EGF homology <EGF1>
 F:240-271/Domain: EGF homology <EGF2>
 F:318-349/Domain: EGF homology <EGF2>
 F:473-504/Domain: EGF homology <EGF3>
 F:853-884/Domain: EGF homology <EGF3>
 F:928-959/Domain: EGF homology <EGF4>
 F:1338-1870/Domain: ankyrin repeat homology <AN1>
 F:1871-1903/Domain: ankyrin repeat homology <AN2>
 F:1905-1937/Domain: ankyrin repeat homology <AN3>
 F:1938-1970/Domain: ankyrin repeat homology <AN4>
 F:1971-2003/Domain: ankyrin repeat homology <AN5>

Query Match 59.7%; Score 83; DB 2; Length 2321;
 Best Local Similarity 41.9%; Pred. No. 0.00037;
 Matches 13; Conservative 2; Mismatches 16; Indels 0; Gaps 0;

QY 1 CDXXXCXKXGKXCDXXCNNAACXXDXDC 31
 Db 1387 CPRAACQAKRGDRCDBECNSPGCGWDGDC 1417

RESULT 13
 S45306
 notch 3 protein - mouse
 C/Species: Mus musculus (house mouse)
 C/Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 02-Aug-2002
 C/Accession: S45306
 R:Lardelli, M.; Dahlstrand, J.; Lendahl, U.
 Mech. Dev. 46, 123-136, 1994
 A>Title: The novel Notch homologue mouse Notch 3 lacks specific epidermal growth factor-
 A/Reference number: S45306; MUID:95001556; PMID:7918097
 A:Accession: S45306
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-2318 <LAR>
 A/Cross-references: EMBL:X74760; NID:g483580; PIDN:CAA52776.1; PID:g483581
 C/Superfamily: notch protein; ankyrin repeat homology; EGF homology
 F:163-195/Domain: EGF homology <EGF1>
 F:474-505/Domain: EGF homology <EGF1>
 F:854-885/Domain: EGF homology <EGF2>
 F:1839-1871/Domain: ankyrin repeat homology <AN1>
 F:1872-1904/Domain: ankyrin repeat homology <AN2>
 F:1906-1938/Domain: ankyrin repeat homology <AN3>
 F:1939-1971/Domain: ankyrin repeat homology <AN4>
 F:1972-2004/Domain: ankyrin repeat homology <AN5>

Query Match 59.0%; Score 82; DB 2; Length 2318;
 Best Local Similarity 41.9%; Pred. No. 0.00051;
 Matches 13; Conservative 1; Mismatches 17; Indels 0; Gaps 0;

QY 1 CDXXXCXKXGKXCDXXCNNAACXXDXDC 31
 Db 1388 CPRAACQAKRGDRCDBECNSPGCGWDGDC 1418

RESULT 14
 A35844
 Notch protein - African clawed frog
 C/Species: Xenopus laevis (African clawed frog)
 C/Date: 12-Oct-1990 #sequence_revision 12-Oct-1990 #text_change 02-Aug-2002

C/Accession: A35844
 R:Coffman, C.; Harris, W.; Kintner, C.
 Science 249, 1438-1441, 1990
 A>Title: Notch, the Xenopus homolog of Drosophila notch.
 A/Reference number: A35844; MUID:90385285; PMID:2402639
 A:Accession: A35844
 A:Status: preliminary; nucleic acid sequence not shown; not compared with conceptual tr
 A:Molecule type: mRNA
 A:Residues: 1-2524 <COF>
 C/Superfamily: notch protein; ankyrin repeat homology; EGF homology
 C/Keywords: transmembrane protein
 F:146-177/Domain: EGF homology <EGF1>
 F:184-215/Domain: EGF homology <EGF1>
 F:222-254/Domain: EGF homology <EGF2>
 F:456-487/Domain: EGF homology <EGF2>
 F:757-788/Domain: EGF homology <EGF3>
 F:1025-1056/Domain: EGF homology <EGF3>
 F:1924-1956/Domain: ankyrin repeat homology <AN1>
 F:1957-1989/Domain: ankyrin repeat homology <AN2>
 F:1991-2023/Domain: ankyrin repeat homology <AN3>
 F:2024-2056/Domain: ankyrin repeat homology <AN4>
 F:2057-2089/Domain: ankyrin repeat homology <AN5>

Query Match 57.6%; Score 80; DB 2; Length 2524;
 Best Local Similarity 41.9%; Pred. No. 0.001;
 Matches 13; Conservative 2; Mismatches 16; Indels 0; Gaps 0;

QY 1 CDXXXCXKXGKXCDXXCNNAACXXDXDC 31
 Db 1448 CENEGCSELDNKNVCNANCNNAACGWDGDC 1478

RESULT 15
 A32901
 glp1 protein precursor - Caenorhabditis elegans
 C/Species: Caenorhabditis elegans
 C/Date: 20-Dec-1989 #sequence_revision 04-Sep-1992 #text_change 20-Sep-1999
 C/Accession: A32901
 R:Yochem, J.; Greenwald, I.
 Cell 58, 553-563, 1989
 A>Title: glp-1 and lin-12, genes implicated in distinct cell-cell interactions in Caenor
 A/Reference number: A32901; MUID:89336787; PMID:2758466
 A:Accession: A32901
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1295 <YOC>
 A/Cross-references: GB:M25580; NID:g156316; PIDN:AA28058.1; PID:g156317
 C/Superfamily: unassigned ankyrin repeat proteins; ankyrin repeat homology; EGF homology
 C/Keywords: transmembrane protein
 F:121-151/Domain: EGF homology <EGF1>
 F:275-307/Domain: EGF homology <EGF1>
 F:373-405/Domain: EGF homology <EGF2>
 F:411-442/Domain: EGF homology <EGF2>
 F:450-478/Domain: EGF homology <EGF3>
 F:961-993/Domain: ankyrin repeat homology <AN1>
 F:1074-1106/Domain: ankyrin repeat homology <AN2>
 F:1107-1139/Domain: ankyrin repeat homology <AN3>

Query Match 56.8%; Score 79; DB 2; Length 1295;
 Best Local Similarity 46.2%; Pred. No. 0.00089;
 Matches 12; Conservative 2; Mismatches 12; Indels 0; Gaps 0;

QY 6 CXKXGKXGKXCDXXCNNAACXXDXDC 31
 Db 542 CADGFANGVCNOCNNECLYDGLDC 567

Search completed: May 19, 2004, 16:00:37
 Job time : 2.29658 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 19, 2004, 15:51:04 ; Search time 0.785805 Seconds
(without alignments)
2054.168 Million cell updates/sec

Title: US-09-903-199-4
Perfect score: 139
Sequence: 1 CDXXXCXXKXGNGCXCDXXCNNAACXXDGXDC 31

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database:

SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	105	75.5	2703	1	NOTC_DROME
2	97	69.8	2437	1	NOTC_BRARE
3	92	66.2	1429	1	L112_CAEEL
4	88	63.3	2470	1	NTC2_MOUSE
5	85	61.2	2531	1	NTC1_MOUSE
6	85	61.2	2531	1	NTC1_MOUSE
7	84	60.4	2471	1	NTC2_HUMAN
8	84	60.4	2471	1	NTC2_HUMAN
9	84	60.4	2556	1	NTC1_HUMAN
10	83	59.7	2319	1	NTC3_RAT
11	83	59.7	2319	1	NTC3_RAT
12	82	59.0	2318	1	NTC3_MOUSE
13	80	57.6	2524	1	NOTC_MOUSE
14	79	56.8	1295	1	GLP1_CAEEL
15	79	56.8	1964	1	NTC4_MOUSE
16	73	52.5	2003	1	NTC4_HUMAN
17	61	43.9	67	1	MTA_SPHGR
18	61	43.9	67	1	MTA_SPHGR
19	60	43.2	1627	1	PAPA_HUMAN
20	53	38.1	68	1	MT_LYTP1
21	50	36.0	64	1	MT_STERNE
22	50	36.0	65	1	MTB_STRPU
23	48.5	34.9	616	1	ECAR_ECHCA
24	47.5	34.2	65	1	MT_PABLI
25	46.5	33.5	571	1	DISJ_BOTUA
26	46	33.1	575	1	TRBM_HUMAN
27	45	33.4	64	1	MTA_STRPU
28	45	33.4	72	1	DISI_CROAT
29	45	33.4	73	1	DISI_CROAT
30	45	33.4	73	1	DISI_CROAT
31	45	33.4	73	1	DISI_CROAT
32	45	32.4	73	1	DISI_CROAT
33	45	32.4	73	1	DISI_CROAT
34	45	32.4	73	1	DISI_CROAT
35	45	32.4	73	1	DISI_CROAT
36	45	32.4	73	1	DISI_CROAT
37	45	32.4	73	1	DISI_CROAT
38	45	32.4	73	1	DISI_CROAT
39	45	32.4	73	1	DISI_CROAT
40	45	32.4	73	1	DISI_CROAT
41	45	32.4	73	1	DISI_CROAT
42	45	32.4	73	1	DISI_CROAT
43	45	32.4	73	1	DISI_CROAT
44	45	32.4	73	1	DISI_CROAT
45	45	32.4	73	1	DISI_CROAT

34	45	32.4	2907	1	FBN2_MOUSE	Q61555 mus musculus
35	45	32.4	2911	1	FBN2_HUMAN	P35556 homo sapien
36	44.5	32.0	788	1	ITB6_HUMAN	P18564 homo sapien
37	44	31.7	84	1	SCX9_CENSC	Q95W69 centuroide
38	44	31.7	186	1	AG13_WHEAT	P10969 tritium ae
39	43.5	31.3	610	1	LEM2_HUMAN	P16681 homo sapien
40	43	30.9	316	1	CH18_POPTA	P16061 populus tri
41	43	30.9	4753	1	LDP_MOUSE	Q04833 caenorhabdi
42	42.5	30.6	873	1	LDP_MOUSE	P8156 mus musculus
43	42.5	30.6	873	1	LDP_MOUSE	P35953 cycloclagus
44	42.5	30.6	873	1	LDP_MOUSE	P8166 rattus norv
45	42.5	30.6	2871	1	FBN1_BOVIN	P98133 bos taurus

ALIGNMENTS

RESULT 1

ID NOTC_DROME STANDARD; PRT; 2703 AA.

AC P07207; O97458; P04154; Q9W478;

DT 01-NOV-1986 (Rel. 03, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 15-MAR-2004 (Rel. 43, Last annotation update)

DE Neogenic locus Notch protein precursor.

GN N OR EG:140611.1 OR EG:163A10.2 OR CG3936.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI_TaxID=7227;

RN (1)

RP SEQUENCE FROM N.A.

RC STRAIN=Oregon-R; TISSUE=Embryo;

RX MEDLINE=86079539; PubMed=3935325;

RA Wharton K.A., Johansen K.M., Xu T., Artavanis-Tsakonas S.;

RT "Nucleotide sequence from the neogenic locus notch implies a gene product that shares homology with proteins containing EGF-like repeats.";

RL Cell 43:567-581(1985).

RN (2)

RP SEQUENCE FROM N.A.

RC STRAIN=Oregon-R; TISSUE=Embryo;

RX MEDLINE=87064624; PubMed=3097517;

RA Kidd S., Kelley M.R., Young M.W.;

RT "Sequence of the notch locus of Drosophila melanogaster: relationship of the encoded protein to mammalian clotting and growth factors.";

RL Mol. Cell. Biol. 6:3094-3108(1986).

RN (3)

RP SEQUENCE FROM N.A.

RC STRAIN=Berkley;

RX MEDLINE=20196006; PubMed=10731132;

RA Adam M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,

RA Amanatides P.G., Scherer S.E., Li P.W., Hopkins R.A., Galante R.F.,

RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,

RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,

RA Branton R.C., Rogers Y.-H.C., Blazej R.G., Chapple M., Pfeiffer B.D.,

RA Wan K.H., Doyle K., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,

RA Abail J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,

RA Bailew R.M., Basu A., Baxendale J., Bayraktoglu L., Beasley E.M.,

RA Beeson K.Y., Benos P.V., Berman B.P., Bandaru D., Bolshakov S.,

RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,

RA Burris K.C., Busam D.A., Butler H., Caden E., Center A., Chandra I.,

RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,

RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,

RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,

RA Dubler K.J., Evangelista C.C., Ferraz C., Ferriere S., Fleischmann W.,

RA Fodor C., Gabriellian A.E., Gary N.S., Gelbart W.M., Glasser K.,

RA Goddek A., Gong F., Gottrell J.H., Gu Z., Guan P., Harris M.,

RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,

RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,

RA Jatali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,

RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,

RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merklov G., Mlishina N.V., Mobarry C., Morris J., Mostreli A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Murry D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusser D.R., Pacle J.M.,
 RA Palazzolo M., Piltman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svitek R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri U.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
 RT "The genome sequence of *Drosophila melanogaster*,"
 RL Science 287:2185-2195(2000).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Oregon-R;
 RX MEDLINE=20196011; PubMed=10731137;
 RA Benos P.V., Galt M.K., Ashburner M., Murphy L., Harris D.,
 RA Barrell B.G., Ferraz C., Vidal S., Brun C., Demallies J., Gadieu E.,
 RA Dreano S., Gloux S., Lelaure V., Motier S., Galibert F., Borkova D.,
 RA Minana B., Kafatos F.C., Louis C., Siden-Kiamos I., Bolshakov S.,
 RA Papagiannakis G., Spanos L., Cox S., Madueno E., de Pablo B.,
 RA Modolell J., Peter A., Schoettler P., Werner M., Mouricot F.,
 RA Belhert N., Dowe G., Schaefer U., Jaekle H., Bucheton A.,
 RA Callister D.M., Campbell L.A., Darlamitsou A., Henderson N.S.,
 RA McMillan P.J., Salles C., Tait E.A., Valenti P., Sanders R.D.C.,
 RA Glover D.M.,
 RT "From sequence to chromosome: the tip of the X chromosome of *D. melanogaster*,"
 RL Science 287:2220-2222(2000).
 RN [5]
 RP SEQUENCE OF 2505-2611 FROM N.A.
 RX MEDLINE=85093329; PubMed=2981631;
 RA Wharton K.A., Yedvobnick B., Finnerty V.G., Artavanis-Tsakonas S.,
 RT "opa: a novel family of transcribed repeats shared by the Notch locus and other developmentally regulated loci in *D. melanogaster*,"
 RL Cell 40:55-62(1985).
 RN [6]
 RP SEQUENCE OF 1-8 FROM N.A.
 RX MEDLINE=87257846; PubMed=3037327;
 RA Kelley M.R., Kidd S., Berg R.L., Young M.W.,
 RT "Restriction of P-element insertions at the Notch locus of *Drosophila melanogaster*,"
 RL Mol. Cell. Biol. 7:1545-1548(1987).
 RN [7]
 RP INTERACTION WITH DX AND MUTANT SU42C.
 RX MEDLINE=94215489; PubMed=8162848;
 RA Diederich R.J., Matsuno K., Hing H., Artavanis-Tsakonas S.,
 RT "Cyclosolic interaction between *deltex* and Notch ankyrin repeats implicates *deltex* in the Notch signaling pathway,"
 RL Development 120:473-481(1994).
 RN [8]
 RP INTERACTION WITH DX.
 RX MEDLINE=95401878; PubMed=7671825;
 RA Matsuno K., Diederich R.J., Go M.D., Blaumel C.M.,
 RT "Deltex acts as a positive regulator of Notch signaling through interactions with the Notch ankyrin repeats,"
 RL Development 121:2633-2644(1995).
 RN [9]
 RP S3 CLEAVAGE BY PSN.
 RX MEDLINE=99221487; PubMed=10206646;
 RA Struhl G., Greenwald I.,
 RT "Presentin is required for activity and nuclear access of Notch in *Drosophila*,"
 RL Nature 398:522-525(1999).
 RN [10]
 RP S3 CLEAVAGE BY PSN.
 RX MEDLINE=99221488; PubMed=10206647;

RA Ye Y., Lukinova N., Fortini M.E.,
 RT "Neurogenic phenotypes and altered Notch processing in *Drosophila*
 RT Presentin mutants,"
 RL Nature 398:525-529(1999).
 RN [11]
 RP S2 CLEAVAGE BY KUZ.
 RX MEDLINE=21657146; PubMed=11799064;
 RA Lieber T., Kidd S., Young M.W.,
 RT "kubandian-mediated cleavage of *Drosophila* Notch,"
 RL Genes Dev. 16:209-221(2002).
 RN [12]
 RP MUTANT MCD5.
 RX MEDLINE=21575956; PubMed=11719214;
 RA Ramain P., Knechtman K., Seugnet L., Artobagast N., Ackermann C.,
 RA Heltzer P.,
 RT "Novel Notch alleles reveal a Deltex-dependent pathway repressing neural fate,"
 RL Curr. Biol. 11:1729-1738(2001).
 RN [13]
 RP REVIEW.
 RX MEDLINE=22256570; PubMed=12369105;
 RA Porfin P.,
 RT "General outlines of the molecular genetics of the Notch signalling pathway in *Drosophila melanogaster*: a review,"
 RL Hereditas 136:89-96(2002).
 CC -1- FUNCTION: Signaling protein, which regulates, with both positive and negative signals, the differentiation of at least central and peripheral nervous system and eye, wing disk, oogenesis, segmental inhibition or induction. Functions as a receptor for membrane-bound ligands Delta and Serrate to regulate cell-fate determination. Upon ligand activation, and releasing from the cell membrane, the Notch intracellular domain (NICD) forms a transcriptional activator complex with Su(H) (Suppressor of hairless) and activates genes of the Ets1 complex. Essential for proper differentiation of ectoderm.
 CC -1- SUBUNIT: Interacts with Su(H) when activated. Interacts with Dlx via its ANK repeats.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Upon activation and S3 cleavage, it is released from the cell membrane and enters into the nucleus in conjunction with Su(H).
 CC -1- PTM: Upon binding its ligands such as Delta or Serrate, it is cleaved (S2 cleavage) in its extracellular domain, close to the transmembrane domain. S2 cleavage is probably mediated by Kuz. It is then cleaved (S3 cleavage) downstream of its transmembrane domain, releasing it from the cell membrane. S3 cleavage requires Psn.
 CC -1- SIMILARITY: Belongs to the NOTCH family.
 CC -1- SIMILARITY: Contains 36 EGF-like domains.
 CC -1- SIMILARITY: Contains 3 Lin/Notch repeats.
 CC -1- SIMILARITY: Contains 6 ANK repeats.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL, M16152; AAB59220.1; -
 DR EMBL, M16153; AAB59220.1; JOINED.
 DR EMBL, M16149; AAB59220.1; JOINED.
 DR EMBL, M16150; AAB59220.1; JOINED.
 DR EMBL, M16151; AAB59220.1; JOINED.
 DR EMBL, M16152; AAB59220.1; -
 DR EMBL, M13689; AAA28725.1; JOINED.
 DR EMBL, M13689; AAA28725.1; JOINED.
 DR EMBL, M13689; AAA28725.1; JOINED.
 DR EMBL, AE003426; AAF45848.2; -
 DR EMBL, AL0035436; CAB37610.1; -
 DR EMBL, AL0035395; CAB37610.1; JOINED.
 DR EMBL, M12175; AAA74496.1; -
 DR EMBL, M16025; AAA28726.1; -

```

Query Match      75.5%; Score 105; DB 1; Length 2703;
Best Local Similarity 54.8%; Pred. No. 7,1e-08;
Matches 17; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

1 CDXXXXKXKXNGCDXXCNNAACXKXDC 31
Db      1482 CDXGCTERKQNGICDSQNTYACNPDGND 1512

RESULT 2
NTCL BRARE      STANDARD; PRT; 2437 AA.
ID NTCL BRARE
AC P46530;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Neurogenic locus notch homolog protein 1 precursor.
GN NOTCH1 OR NOTCH.
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxId=7955;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RX MEDLINE=94128602; PubMed=8297791;
RA Bierkamp C., Campos-Ortega J.A.;
RT "A zebrafish homologue of the Drosophila neurogenic gene Notch and
RL its pattern of transcription during early embryogenesis.";
Mech. Dev. 43:87-100(1993).
CC -1- FUNCTION: Implicated in cell fate specifications during
CC embryo development. May be involved in the formation of the
CC neural plate, notochord and brain vesicles.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- DEVELOPMENTAL STAGE: Expressed in all cells in pregastrulation
CC stages. During gastrulation is differentially expressed,
CC accumulating predominantly in the prechordal mesoderm and
CC notochord. At the end of gastrulation, expressed along the
CC anterior-posterior axis including the developing neural plate
CC and differentiating mesoderm. Also present in the developing
CC brain and head regions.
CC -1- SIMILARITY: Belongs to the NOTCH family.
CC -1- SIMILARITY: Contains 36 EGF-like domains.
CC -1- SIMILARITY: Contains 3 Lin/Notch repeats.
CC -1- SIMILARITY: Contains 6 ANK repeats.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
DR EMBL; X69088; CAA48831.1; -
DR PIR; S42612; S42612.
DR HSSP; P00740; IEDM.
DR ZFIN; ZDB-GENE-990415-173; notch1a.
DR InterPro; IPR002110; ANK.
DR InterPro; IPR000152; Asx_hydroxyl_S.
DR InterPro; IPR001861; EGF_Ca.
DR InterPro; IPR001438; EGF_II.
DR InterPro; IPR006209; EGF_like.
DR InterPro; IPR002049; laminin_EGF.
DR InterPro; IPR008297; Notch.
DR InterPro; IPR000800; Notch_dom.
DR Pfam; PF00023; ank; 6.
DR Pfam; PF00008; EGF; 36.
DR Pfam; PF00066; notch; 3.
DR PIRSF; PIRSF002279; Notch; 1.

```

```

DR PRINTS; PR00010; EGFRLD.
DR PRINTS; PR00011; EGFANININ.
DR PRINTS; PR01452; NOTCH.
DR SMART; SM00248; ANK; 6.
DR SMART; SM00179; EGF_Ca; 22.
DR SMART; SM00004; NL; 3.
DR PROSITE; PS50297; ANK_REPEAT; 4.
DR PROSITE; PS50088; ASX_HYDROXYL; 23.
DR PROSITE; PS00010; ASX_HYDROXYL; 23.
DR PROSITE; PS00022; EGF_1; 34.
DR PROSITE; PS01186; EGF_2; 28.
DR PROSITE; PS50026; EGF_3; 36.
DR PROSITE; PS01187; EGF_Ca; 22.
KW Receptor; Transcription regulation; Activator; Differentiation;
KW Developmental protein; Neutogenesis; Repeat; ANK repeat;
KW EGF-like domain; Transmembrane; Glycoprotein; Signal.
FT SIGNAL 1 20
FT CHAIN 21 2437
FT DOMAIN 21 1724
FT TRANSMEM 1725 1747
FT DOMAIN 1748 2437
FT DOMAIN 21 57
FT DOMAIN 58 98
FT DOMAIN 101 138
FT DOMAIN 139 175
FT DOMAIN 177 215
FT DOMAIN 217 254
FT DOMAIN 256 292
FT DOMAIN 294 332
FT DOMAIN 334 370
FT DOMAIN 371 409
FT DOMAIN 411 449
FT DOMAIN 451 487
FT DOMAIN 489 524
FT DOMAIN 526 562
FT DOMAIN 564 599
FT DOMAIN 601 637
FT DOMAIN 639 674
FT DOMAIN 676 712
FT DOMAIN 714 749
FT DOMAIN 751 787
FT DOMAIN 789 825
FT DOMAIN 827 865
FT DOMAIN 867 903
FT DOMAIN 905 941
FT DOMAIN 943 979
FT DOMAIN 981 1017
FT DOMAIN 1019 1055
FT DOMAIN 1057 1093
FT DOMAIN 1095 1141
FT DOMAIN 1143 1179
FT DOMAIN 1181 1217
FT DOMAIN 1219 1263
FT DOMAIN 1265 1303
FT DOMAIN 1305 1344
FT DOMAIN 1346 1382
FT DOMAIN 1385 1423
FT REPEAT 1426 1466
FT REPEAT 1468 1506
FT REPEAT 1521 1561
FT REPEAT 1567 1607
FT REPEAT 1615 1654
FT REPEAT 1657 1697
FT REPEAT 1698 1738
FT REPEAT 1739 1779
FT REPEAT 1780 1820
FT REPEAT 1821 1861
FT REPEAT 1862 1902
FT REPEAT 1903 1943
FT REPEAT 1944 1984
FT REPEAT 1985 2025
FT REPEAT 2026 2066
FT REPEAT 2067 2107
FT REPEAT 2108 2148
FT REPEAT 2149 2189
FT REPEAT 2190 2230
FT REPEAT 2231 2271
FT REPEAT 2272 2312
FT REPEAT 2313 2353
FT REPEAT 2354 2394
FT REPEAT 2395 2435
FT REPEAT 2436 2476
FT REPEAT 2477 2517
FT REPEAT 2518 2558
FT REPEAT 2559 2599
FT REPEAT 2600 2640
FT REPEAT 2641 2681
FT REPEAT 2682 2722
FT REPEAT 2723 2763
FT REPEAT 2764 2804
FT REPEAT 2805 2845
FT REPEAT 2846 2886
FT REPEAT 2887 2927
FT REPEAT 2928 2968
FT REPEAT 2969 3009
FT REPEAT 3010 3050
FT REPEAT 3051 3091
FT REPEAT 3092 3132
FT REPEAT 3133 3173
FT REPEAT 3174 3214
FT REPEAT 3215 3255
FT REPEAT 3256 3296
FT REPEAT 3297 3337
FT REPEAT 3338 3378
FT REPEAT 3379 3419
FT REPEAT 3420 3460
FT REPEAT 3461 3501
FT REPEAT 3502 3542
FT REPEAT 3543 3583
FT REPEAT 3584 3624
FT REPEAT 3625 3665
FT REPEAT 3666 3706
FT REPEAT 3707 3747
FT REPEAT 3748 3788
FT REPEAT 3789 3829
FT REPEAT 3830 3870
FT REPEAT 3871 3911
FT REPEAT 3912 3952
FT REPEAT 3953 3993
FT REPEAT 3994 4034
FT REPEAT 4035 4075
FT REPEAT 4076 4116
FT REPEAT 4117 4157
FT REPEAT 4158 4198
FT REPEAT 4199 4239
FT REPEAT 4240 4280
FT REPEAT 4281 4321
FT REPEAT 4322 4362
FT REPEAT 4363 4403
FT REPEAT 4404 4444
FT REPEAT 4445 4485
FT REPEAT 4486 4526
FT REPEAT 4527 4567
FT REPEAT 4568 4608
FT REPEAT 4609 4649
FT REPEAT 4650 4690
FT REPEAT 4691 4731
FT REPEAT 4732 4772
FT REPEAT 4773 4813
FT REPEAT 4814 4854
FT REPEAT 4855 4895
FT REPEAT 4896 4936
FT REPEAT 4937 4977
FT REPEAT 4978 5018
FT REPEAT 5019 5059
FT REPEAT 5060 5100
FT REPEAT 5101 5141
FT REPEAT 5142 5182
FT REPEAT 5183 5223
FT REPEAT 5224 5264
FT REPEAT 5265 5305
FT REPEAT 5306 5346
FT REPEAT 5347 5387
FT REPEAT 5388 5428
FT REPEAT 5429 5469
FT REPEAT 5470 5510
FT REPEAT 5511 5551
FT REPEAT 5552 5592
FT REPEAT 5593 5633
FT REPEAT 5634 5674
FT REPEAT 5675 5715
FT REPEAT 5716 5756
FT REPEAT 5757 5797
FT REPEAT 5798 5838
FT REPEAT 5839 5879
FT REPEAT 5880 5920
FT REPEAT 5921 5961
FT REPEAT 5962 6002
FT REPEAT 6003 6043
FT REPEAT 6044 6084
FT REPEAT 6085 6125
FT REPEAT 6126 6166
FT REPEAT 6167 6207
FT REPEAT 6208 6248
FT REPEAT 6249 6289
FT REPEAT 6290 6330
FT REPEAT 6331 6371
FT REPEAT 6372 6412
FT REPEAT 6413 6453
FT REPEAT 6454 6494
FT REPEAT 6495 6535
FT REPEAT 6536 6576
FT REPEAT 6577 6617
FT REPEAT 6618 6658
FT REPEAT 6659 6699
FT REPEAT 6700 6740
FT REPEAT 6741 6781
FT REPEAT 6782 6822
FT REPEAT 6823 6863
FT REPEAT 6864 6904
FT REPEAT 6905 6945
FT REPEAT 6946 6986
FT REPEAT 6987 7027
FT REPEAT 7028 7068
FT REPEAT 7069 7109
FT REPEAT 7110 7150
FT REPEAT 7151 7191
FT REPEAT 7192 7232
FT REPEAT 7233 7273
FT REPEAT 7274 7314
FT REPEAT 7315 7355
FT REPEAT 7356 7396
FT REPEAT 7397 7437
FT REPEAT 7438 7478
FT REPEAT 7479 7519
FT REPEAT 7520 7560
FT REPEAT 7561 7601
FT REPEAT 7602 7642
FT REPEAT 7643 7683
FT REPEAT 7684 7724
FT REPEAT 7725 7765
FT REPEAT 7766 7806
FT REPEAT 7807 7847
FT REPEAT 7848 7888
FT REPEAT 7889 7929
FT REPEAT 7930 7970
FT REPEAT 7971 8011
FT REPEAT 8012 8052
FT REPEAT 8053 8093
FT REPEAT 8094 8134
FT REPEAT 8135 8175
FT REPEAT 8176 8216
FT REPEAT 8217 8257
FT REPEAT 8258 8298
FT REPEAT 8299 8339
FT REPEAT 8340 8380
FT REPEAT 8381 8421
FT REPEAT 8422 8462
FT REPEAT 8463 8503
FT REPEAT 8504 8544
FT REPEAT 8545 8585
FT REPEAT 8586 8626
FT REPEAT 8627 8667
FT REPEAT 8668 8708
FT REPEAT 8709 8749
FT REPEAT 8750 8790
FT REPEAT 8791 8831
FT REPEAT 8832 8872
FT REPEAT 8873 8913
FT REPEAT 8914 8954
FT REPEAT 8955 8995
FT REPEAT 8996 9036
FT REPEAT 9037 9077
FT REPEAT 9078 9118
FT REPEAT 9119 9159
FT REPEAT 9160 9200
FT REPEAT 9201 9241
FT REPEAT 9242 9282
FT REPEAT 9283 9323
FT REPEAT 9324 9364
FT REPEAT 9365 9405
FT REPEAT 9406 9446
FT REPEAT 9447 9487
FT REPEAT 9488 9528
FT REPEAT 9529 9569
FT REPEAT 9570 9610
FT REPEAT 9611 9651
FT REPEAT 9652 9692
FT REPEAT 9693 9733
FT REPEAT 9734 9774
FT REPEAT 9775 9815
FT REPEAT 9816 9856
FT REPEAT 9857 9897
FT REPEAT 9898 9938
FT REPEAT 9939 9979
FT REPEAT 9980 10020
FT REPEAT 10021 10061
FT REPEAT 10062 10102
FT REPEAT 10103 10143
FT REPEAT 10144 10184
FT REPEAT 10185 10225
FT REPEAT 10226 10266
FT REPEAT 10267 10307
FT REPEAT 10308 10348
FT REPEAT 10349 10389
FT REPEAT 10390 10430
FT REPEAT 10431 10471
FT REPEAT 10472 10512
FT REPEAT 10513 10553
FT REPEAT 10554 10594
FT REPEAT 10595 10635
FT REPEAT 10636 10676
FT REPEAT 10677 10717
FT REPEAT 10718 10758
FT REPEAT 10759 10799
FT REPEAT 10800 10840
FT REPEAT 10841 10881
FT REPEAT 10882 10922
FT REPEAT 10923 10963
FT REPEAT 10964 11004
FT REPEAT 11005 11045
FT REPEAT 11046 11086
FT REPEAT 11087 11127
FT REPEAT 11128 11168
FT REPEAT 11169 11209
FT REPEAT 11210 11250
FT REPEAT 11251 11291
FT REPEAT 11292 11332
FT REPEAT 11333 11373
FT REPEAT 11374 11414
FT REPEAT 11415 11455
FT REPEAT 11456 11496
FT REPEAT 11497 11537
FT REPEAT 11538 11578
FT REPEAT 11579 11619
FT REPEAT 11620 11660
FT REPEAT 11661 11701
FT REPEAT 11702 11742
FT REPEAT 11743 11783
FT REPEAT 11784 11824
FT REPEAT 11825 11865
FT REPEAT 11866 11906
FT REPEAT 11907 11947
FT REPEAT 11948 11988
FT REPEAT 11989 12029
FT REPEAT 12030 12070
FT REPEAT 12071 12111
FT REPEAT 12112 12152
FT REPEAT 12153 12193
FT REPEAT 12194 12234
FT REPEAT 12235 12275
FT REPEAT 12276 12316
FT REPEAT 12317 12357
FT REPEAT 12358 12398
FT REPEAT 12399 12439
FT REPEAT 12440 12480
FT REPEAT 12481 12521
FT REPEAT 12522 12562
FT REPEAT 12563 12603
FT REPEAT 12604 12644
FT REPEAT 12645 12685
FT REPEAT 12686 12726
FT REPEAT 12727 12767
FT REPEAT 12768 12808
FT REPEAT 12809 12849
FT REPEAT 12850 12890
FT REPEAT 12891 12931
FT REPEAT 12932 12972
FT REPEAT 12973 13013
FT REPEAT 13014 13054
FT REPEAT 13055 13095
FT REPEAT 13096 13136
FT REPEAT 13137 13177
FT REPEAT 13178 13218
FT REPEAT 13219 13259
FT REPEAT 13260 13300
FT REPEAT 13301 13341
FT REPEAT 13342 13382
FT REPEAT 13383 13423
FT REPEAT 13424 13464
FT REPEAT 13465 13505
FT REPEAT 13506 13546
FT REPEAT 13547 13587
FT REPEAT 13588 13628
FT REPEAT 13629 13669
FT REPEAT 13670 13710
FT REPEAT 13711 13751
FT REPEAT 13752 13792
FT REPEAT 13793 13833
FT REPEAT 13834 13874
FT REPEAT 13875 13914
FT REPEAT 13915 13955
FT REPEAT 13956 13996
FT REPEAT 13997 14036
FT REPEAT 14037 14077
FT REPEAT 14078 14118
FT REPEAT 14119 14159
FT REPEAT 14160 14200
FT REPEAT 14201 14241
FT REPEAT 14242 14282
FT REPEAT 14283 14323
FT REPEAT 14324 14364
FT REPEAT 14365 14405
FT REPEAT 14406 14446
FT REPEAT 14447 14487
FT REPEAT 14488 14528
FT REPEAT 14529 14569
FT REPEAT 14570 14610
FT REPEAT 14611 14651
FT REPEAT 14652 14692
FT REPEAT 14693 14733
FT REPEAT 14734 14774
FT REPEAT 14775 14814
FT REPEAT 14815 14855
FT REPEAT 14856 14896
FT REPEAT 14897 14936
FT REPEAT 14937 14977
FT REPEAT 14978 15017
FT REPEAT 15018 15058
FT REPEAT 15059 15099
FT REPEAT 15100 15140
FT REPEAT 15141 15181
FT REPEAT 15182 15222
FT REPEAT 15223 15263
FT REPEAT 15264 15304
FT REPEAT 15305 15345
FT REPEAT 15346 15386
FT REPEAT 15387 15427
FT REPEAT 15428 15468
FT REPEAT 15469 15509
FT REPEAT 15510 15550
FT REPEAT 15551 15591
FT REPEAT 15592 15632
FT REPEAT 15633 15673
FT REPEAT 15674 15714
FT REPEAT 15715 15755
FT REPEAT 15756 15796
FT REPEAT 15797 15837
FT REPEAT 15838 15878
FT REPEAT 15879 15919
FT REPEAT 15920 15960
FT REPEAT 15961 16001
FT REPEAT 16002 16042
FT REPEAT 16043 16083
FT REPEAT 16084 16124
FT REPEAT 16125 16165
FT REPEAT 16166 16206
FT REPEAT 16207 16247
FT REPEAT 16248 16288
FT REPEAT 16289 16329
FT REPEAT 16330 16370
FT REPEAT 16371 16411
FT REPEAT 16412 16452
FT REPEAT 16453 16493
FT REPEAT 16494 16534
FT REPEAT 16535 16575
FT REPEAT 16576 16616
FT REPEAT 16617 16657
FT REPEAT 16658 16698
FT REPEAT 16699 16739
FT REPEAT 16740 16780
FT REPEAT 16781 16821
FT REPEAT 16822 16862
FT REPEAT 16863 16903
FT REPEAT 16904 16944
FT REPEAT 16945 16985
FT REPEAT 16986 17026
FT REPEAT 17027 17067
FT REPEAT 17068 17108
FT REPEAT 17109 17149
FT REPEAT 17150 17190
FT REPEAT 17191 17231
FT REPEAT 17232 17272
FT REPEAT 17273 17313
FT REPEAT 17314 17354
FT REPEAT 17355 17395
FT REPEAT 17396 17436
FT REPEAT 17437 17477
FT REPEAT 17478 17518
FT REPEAT 17519 17559
FT REPEAT 17560 17600
FT REPEAT 17601 17641
FT REPEAT 17642 17682
FT REPEAT 17683 17723
FT REPEAT 17724 17764
FT REPEAT 17765 17805
FT REPEAT 17806 17846
FT REPEAT 17847 17887
FT REPEAT 17888 17928
FT REPEAT 17929 17969
FT REPEAT 17970 18010
FT REPEAT 18011 18051
FT REPEAT 18052 18092
FT REPEAT 18093 18133
FT REPEAT 18134 18174
FT REPEAT 18175 18215
FT REPEAT 18216 18256
FT REPEAT 18257 18297
FT REPEAT 18298 18338
FT REPEAT 18339 18379
FT REPEAT 18380 18420
FT REPEAT 18421 18461
FT REPEAT 18462 18502
FT REPEAT 18503 18543
FT REPEAT 18544 18584
FT REPEAT 18585 18625
FT REPEAT 18626 18666
FT REPEAT 18667 18707
FT REPEAT 18708 18748
FT REPEAT 18749 18789
FT REPEAT 18790 18830
FT REPEAT 18831 18871
FT REPEAT 18872 18912
FT REPEAT 18913 18953
FT REPEAT 18954 18994
FT REPEAT 18995 19035
FT REPEAT 19036 19075
FT REPEAT 19076 19116
FT REPEAT 19117 19157
FT REPEAT 19158 19198
FT REPEAT 19199 19239
FT REPEAT 19240 19280
FT REPEAT 19281 19321
FT REPEAT 19322 19362
FT REPEAT 19363 19403
FT REPEAT 19404 19444
FT REPEAT 19445 19485
FT REPEAT 19486 19526
FT REPEAT 19527 19567
FT REPEAT 19568 19608
FT REPEAT 19609 19649
FT REPEAT 19650 19690
FT REPEAT 19691 19731
FT REPEAT 19732 19772
FT REPEAT 19773 19813
FT REPEAT 19814 19854
FT REPEAT 19855 19895
FT REPEAT 19896 19936
FT REPEAT 19937 19977
FT REPEAT 19978 20018
FT REPEAT 20019 20059
FT REPEAT 20060 20100
FT REPEAT 20101 20141
FT REPEAT 20142 20182
FT REPEAT 20183 20223
FT REPEAT 20224 20264
FT REPEAT 20265 20305
FT REPEAT 20306 20346
FT REPEAT 20347 20387
FT REPEAT 20388 20428
FT REPEAT 20429 20469
FT REPEAT 20470 20510
FT REPEAT 20511 20551
FT REPEAT 20552 20592
FT REPEAT 20593 20633
FT REPEAT 20634 20674
FT REPEAT 20675 20715
FT REPEAT 20716 20756
FT REPEAT 20757 20797
FT REPEAT 20798 20838
FT REPEAT 20839 20879
FT REPEAT 20880 20920
FT REPEAT 20921 20961
FT REPEAT 20962 21002
FT REPEAT 21003 21043
FT REPEAT 21044 21084
FT REPEAT 21085 21125
FT REPEAT 21126 21166
FT REPEAT 21167 21207
FT REPEAT 21208 21248
FT REPEAT 21249 21289
FT REPEAT 21290 21330
FT REPEAT 21331 21371
FT REPEAT 21372 21412
FT REPEAT 21413 21453
FT REPEAT 21454 21494
FT REPEAT 21495 21535
FT REPEAT 21536 21576
FT REPEAT 21577 21617
FT REPEAT 21618 21658
FT REPEAT 21659 21700
FT REPEAT 21701 21741
FT REPEAT 21742 21782
FT REPEAT 21783 21823
FT REPEAT 21824 21864
FT REPEAT 21865 21905
FT REPEAT 21906 21946
FT REPEAT 21947 21987
FT REPEAT 21988 22028
FT REPEAT 22029 22069
FT REPEAT 22070 22110
FT REPEAT 22111 22151
FT REPEAT 22152 22192
FT REPEAT 22193 22233
FT REPEAT 22234 22274
FT REPEAT 22275 22315
FT REPEAT 22316 22356
FT REPEAT 22357 22397
FT REPEAT 22398 22438
FT REPEAT 22439 22479
FT REPEAT 22480 22520
FT REPEAT 22521 22561
FT REPEAT 22562 22602
FT REPEAT 22603 22643
FT REPEAT 22644 22684
FT REPEAT 22685 22725
FT REPEAT 22726 22766
FT REPEAT 22767 22807
FT REPEAT 22808 22848
FT REPEAT 22849 22889
FT REPEAT 22890 22930
FT REPEAT 22931 22971
FT REPEAT 22972 23012
FT REPEAT 23013 23053
FT REPEAT 23054 23094
FT REPEAT 23095 23135
FT REPEAT 23136 23176
FT REPEAT 23177 23217
FT REPEAT 23218 23258
FT REPEAT 23259 23299
FT REPEAT 23300 23340
FT REPEAT 23341 23381
FT REPEAT 23382 23422
FT REPEAT 23423 23463
FT REPEAT 23464 23504
FT REPEAT 23505 23545
FT REPEAT 23546 23586
FT REPEAT 23587 23627
FT REPEAT 23628 23668
FT REPEAT 23669 23709
FT REPEAT 23710 23750
FT REPEAT 23751 23791
FT REPEAT 23792 23832
FT REPEAT 23833 23873
FT REPEAT 23874 23914
FT REPEAT 23915 23955
FT REPEAT 23956 23996
FT REPEAT 23997 24037
FT REPEAT 24038 24078
FT REPEAT 24079 24119
FT REPEAT 24120 24160
FT REPEAT 24161 24201
FT REPEAT 24202 24242
FT REPEAT 24243 24283
FT REPEAT 24284 24324
FT REPEAT 24325 24365
FT REPEAT 24366 24406
FT REPEAT 24407 24447
FT REPEAT 24448 24488
FT REPEAT 24489 24529
FT REPEAT 24530 24570
FT REPEAT 24571 24611
FT REPEAT 24612 24652
FT REPEAT 24653 24693
FT REPEAT 24694 24734
FT REPEAT 24735 24775
FT REPEAT 24776 24816
FT REPEAT 24817 24857
FT REPEAT 24858 24898
FT REPEAT 24899 24939
FT REPEAT 24940 24980
FT REPEAT 24981 25021
FT REPEAT 25022 25062
FT REPEAT 25063 25103
FT REPEAT 25104 25144
FT REPEAT 25145 25185
FT REPEAT 25186 25226
FT REPEAT 25227 25267
FT REPEAT 25268 25308
FT REPEAT 25309 25349
FT REPEAT 25350 25390
FT REPEAT 25391 25431
FT REPEAT 25432 25472
FT REPEAT 25473 25513
FT REPEAT 25514 25554
FT REPEAT 25555 25595
FT REPEAT 25596 25636
FT REPEAT 25637 25677
FT REPEAT 25678 25718
FT REPEAT 25719 25759
FT REPEAT 25760 25799
FT REPEAT 25800 25840
FT REPEAT 25841 25881
FT REPEAT 25882 25922
FT REPEAT 25923 25963
FT REPEAT 25964 26004
FT REPEAT 26005 26045
FT REPEAT 26046 26086
FT REPEAT 26087 26126
FT REPEAT 26127 26167
FT REPEAT 26168 26208
FT REPEAT 26209 26249
FT REPEAT 26250 26289
FT REPEAT 26290 26330
FT REPEAT 26331 26371
FT REPEAT 26372 26412
FT REPEAT 26413 26453
FT REPEAT 26454 26494
FT REPEAT 26495 26535
FT REPEAT 26536 26576
FT REPEAT 26577 26617
FT REPEAT 26618 26658
FT REPEAT 26659 26699
FT REPEAT 26700 26740
FT REPEAT 26741 26781
FT REPEAT 26782 26822
FT REPEAT 26823 26863
FT REPEAT 26864 26904
FT REPEAT 26905 26945
FT REPEAT 26946 26986
FT REPEAT 26987 27027
FT REPEAT 27028 27068
FT REPEAT 27069 27109
FT REPEAT 27110 27150
FT REPEAT 27151 27191
FT REPEAT 27192 27232
FT REPEAT 27233 27273
FT REPEAT 27274 27314
FT REPEAT 27315 27355
FT REPEAT 27356 27396
FT REPEAT 27397 27437
FT REPEAT 27438 27478
FT REPEAT 27479 27519
FT REPEAT 27520 27560
FT REPEAT 27561 27601
FT REPEAT 27602 27642
FT REPEAT 27643 27683
FT REPEAT 27684 27724
FT REPEAT 27725 27765
FT REPEAT 27766 27806
FT REPEAT 27807 27847
FT REPEAT 27848 27888
FT REPEAT 27889 27929
FT REPEAT 27930 27970
FT REPEAT 27971 28011
FT REPEAT 28012 28052
FT REPEAT 28053 28093
FT REPEAT 28094 28134
FT REPEAT 28135 28175
FT REPEAT 28176 28216
FT REPEAT 28217 28257
FT REPEAT 28258 28298
FT REPEAT 28299 28339
FT REPEAT 28340 28380
FT REPEAT 28381 28421
FT REPEAT 28422 28462
FT REPEAT 28463 28503
FT REPEAT 28504 28544
FT REPEAT 28545 28585
FT REPEAT 28586 28626
FT REPEAT 28627 28667
FT REPEAT 28668 28708
FT REPEAT 28709 28749
FT REPEAT 28750 28790

```



```
FT DOMAIN 1185 1221 BGF-LIKE 31, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1223 1260 BGF-LIKE 32, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1262 1300 BGF-LIKE 33.
FT DOMAIN 1302 1345 BGF-LIKE 34.
FT DOMAIN 1372 1410 BGF-LIKE 35.
FT REPEAT 1418 1454 LTN/NOTCH 1.
FT REPEAT 1501 1533 LTN/NOTCH 2.
FT REPEAT 1825 1869 ANK 1.

Query Match 63.3% Score 88; DB 1; Length 2470;
Best Local Similarity 45.2%; Pred No. 2.1e-05;
Matches 14; Conservative 2; Mismatches 15; Indels 0; Gaps 0;

OY 1 CDXXXCXXKXGNGXCDXXCNNAACXXDXDC 31
Db 1423 COSGCADKARDGICDBACNSHACQMDGDC 1453

RESULT 5
NTC1 MOUSE STANDARD; PRT; 2531 AA.
AC NT01705; Q06007; Q61905; Q99JC2; Q9QW58; Q9R0X7;
AD 01-NOV-1995 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Neutrogenic locus notch homolog protein 1 precursor (Notch 1) (Notch A)
DE (mRNA) (p300).
GN NOTCH1 OR NOTCH.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCBI_TaxID=10090;
RN RP SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=Embryo;
RX MEDLINE=93194170; PubMed=8449489;
RA Franco del Amo F., Gendron-Maguire M., Swiatek P.J., Jenkins N.A.,
RA Copeland N.G., Gridley T.;
RT "Cloning, analysis, and chromosomal localization of Notch-1, a mouse
RL homolog of Drosophila Notch.";
RN Genomics 15:259-264(1993).
RN [2]
RN RP SEQUENCE OF 731-1899 FROM N.A. (ISOFORM 2), AND DEVELOPMENTAL STAGE.
RC STRAIN=CD-1; TISSUE=Embryo;
RX MEDLINE=93050801; PubMed=1426644;
RA Reame A.G., Conlon R.A., Zimngibl R., Yamaguchi T.P., Rosant J.;
RT "Expression analysis of a Notch homologue in the mouse embryo.";
RN Dev. Biol. 154:377-387(1992).
RN [3]
RN RP SEQUENCE OF 1551-1647 FROM N.A. (ISOFORM 1), AND DEVELOPMENTAL STAGE.
RC TISSUE=Embryo;
RX MEDLINE=93048835; PubMed=1425352;
RA Franco del Amo F., Smith D.E., Swiatek P.J., Gendron-Maguire M.,
RA Greenspan R.J., McMahon A.P., Gridley T.;
RT "Expression pattern of Notch, a mouse homolog of Drosophila Notch,
RN suggests an important role in early postimplantation mouse
RL development.";
RN Development 115:737-744(1992).
RN [4]
RN RP SEQUENCE OF 1161-1547 FROM N.A.
RC STRAIN=C57BL/6 X CBA; TISSUE=Embryo;
RX MEDLINE=93178563; PubMed=8440332;
RA Lardelli M., Lendahl U.;
RT "Notch A and Notch B-two mouse Notch homologues coexpressed in a
RN wide variety of tissues.";
RN Exp. Cell Res. 204:364-372(1993).
RN [5]
RN RP SEQUENCE OF 1659-1673 FROM N.A.
RX MEDLINE=99364499; PubMed=10437788;
RA Lee J.S., Ichimoto A., Yanagawa S.I.;
RT "Murine leukemia provirus-mediated activation of the Notch1 gene leads
RN to induction of HES-1 in a mouse T lymphoma cell line, DL-3.";
RN FBSB Lett. 453:276-280(1999).
```

```
RN [6]
RP SEQUENCE OF 1950-2201 FROM N.A.
RX MEDLINE=98029496; PubMed=9384671;
RA Messierie M., Folio M., Nenis M., Eggert H., Boehm T.;
RT "Dynamic changes in gene expression during in vitro differentiation of
RN mouse embryonic stem cells.";
RN Cytokines Cell. Mol. Ther. 1:139-143(1995).
RN [7]
RP SEQUENCE OF 1655-1659, CLEAVAGE BY FURIN-LIKE CONVERTASE, AND
RX MUTAGENESIS OF 1651-ARG-ARG-1654.
RX MEDLINE=98318619; PubMed=9653148;
RA Logeat F., Bessia C., Brou C., LeBail O., Jarriault S., Seidah N.G.,
RA Israel A.;
RT "The Notch1 receptor is cleaved constitutively by a furin-like
RN convertase.";
RN Proc. Natl. Acad. Sci. U.S.A. 95:8108-8112(1998).
RN [8]
RP PARTIAL SEQUENCE, AND POST-TRANSLATIONAL PROCESSING.
RX MEDLINE=21523956; PubMed=11518718;
RA Saxena M.T., Schroeter E.H., Mumm J.S., Kopan R.;
RT "Murine notch homologs (N1-4) undergo presenilin-dependent
RN proteolysis.";
RN J. Biol. Chem. 276:40268-40273(2001).
RN [9]
RP POST-TRANSLATIONAL PROCESSING.
RX MEDLINE=21374376; PubMed=11459941;
RA Mzuran T., Taniguchi Y., Aoki T., Hashimoto N., Horjo T.;
RT "Conservation of the biochemical mechanisms of signal transduction
RN among mammalian Notch family members.";
RN Proc. Natl. Acad. Sci. U.S.A. 98:9026-9031(2001).
RN [10]
RP INTERACTION WITH DTX1 AND DTX2.
RX MEDLINE=21123790; PubMed=11226752;
RA Kishi N., Tang Z., Maeda Y., Hirai A., Mo R., Ito M., Suzuki S.,
RA Nakao K., Kinoshita T., Kadesch T., Hui C.-C., Attavanis-Teakonas S.,
RA Okano H., Matsuno K.;
RT "Murine homologs of delta define a novel gene family involved in
RN vertebrate Notch signaling and neurogenesis.";
RN Int. J. Dev. Neurosci. 19:21-35(2001).
RN [11]
RP FUNCTION: Functions as a receptor for membrane-bound ligands
CC Jagged1, Jagged2 and Delta1 to regulate cell-fate determination.
CC Upon ligand activation through the released notch intracellular
CC domain (NICD) it forms a transcriptional activator complex with
CC RBP-J kappa and activates genes of the enhancer of split locus.
CC Affects the implementation of differentiation, proliferation and
CC apoptotic programs (by similarity). May play an essential role in
CC postimplantation development, probably in some aspect of cell
CC specification and/or differentiation. May be involved in mesoderm
CC development, somitome formation and neurogenesis. Involved in the
CC maturation of both CD4+ and CD8+ cells in the thymus.
CC [12]
RP SUBUNIT: Heterodimer of a C-terminal fragment N(TW) and a N-
CC terminal fragment N(EC) which are probably linked by disulfide
CC bonds. Interacts with DTX1 and DTX2.
CC [13]
RP SUBCELLULAR LOCATION: Type I membrane protein. Following
CC proteolytical processing NICD is translocated to the nucleus.
CC [14]
RP ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=1;
CC IsoId=Q01705-1; Sequence=Displayed;
CC Name=2;
CC IsoId=Q01705-2; Sequence=VSP_001402, VSP_001403, VSP_001404;
CC Note=No experimental confirmation available;
CC [15]
RP TISSUE SPECIFICITY: Highly expressed in the brain, lung and
CC thymus. Expressed at lower levels in the spleen, bone-marrow,
CC spinal cord, eyes, mammary gland, liver, intestine, skeletal
CC muscle, kidney and heart.
CC [16]
RP DEVELOPMENTAL STAGE: First detected in the mesoderm at 7.5 dpc. By
CC 8.5 dpc highly expressed in presomitic mesoderm, mesenchyme and
CC endothelial cells, while much lower levels are seen in the
CC neuroepithelium. Between 9.5-10.5 dpc expressed at high levels in
CC the neuroepithelium. At 13.5 dpc expressed in the surface
CC ectoderm, eye and developing whisker follicles.
CC [17]
RP PTM: Synthesized in the endoplasmic reticulum as an inactive form
```


CC which is proteolytically cleaved by a furin-like convertase in the
 CC trans-Golgi network before it reaches the plasma membrane to yield
 CC an active, ligand-accessible form. Cleavage results in a C-
 CC terminal fragment N(TM) and a N-terminal fragment N(EC). Following
 CC ligand binding, it is cleaved by TNF-alpha converting enzyme
 CC (TACE) to yield a membrane-associated intermediate fragment called
 CC notch extracellular truncation (NEXTR). This fragment is then
 CC cleaved by presenilin dependent gamma-secretase to release a
 CC notch-derived peptide containing the intracellular domain (NICD)
 CC from the membrane.
 CC -1- PTM: Phosphorylated.
 CC -1- SIMILARITY: Belongs to the NOTCH family.
 CC -1- SIMILARITY: Contains 36 EGF-like domains.
 CC -1- SIMILARITY: Contains 3 Lin/Notch repeats.
 CC -1- SIMILARITY: Contains 5 ANK repeats.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: Z1886; CAA77941.1; -
 CC EMBL: L02613; AAK14898.1; -
 CC EMBL: X68278; CAA48359.1; -
 CC EMBL: AJ238029; CAB40733.1; -
 CC EMBL: X82562; CAA57909.1; -
 CC PIR: A46019; A46019.
 CC PIR: B49175; B49175.
 CC HSSP: P00740; 1EDM.
 CC MGD: MGI:97363; Notch1.
 CC GO: GO:0005887; C:integral to plasma membrane; IC.
 CC GO: GO:0005515; P:protein binding; IPT.
 CC GO: GO:0030154; P:cell differentiation; IMP.
 CC GO: GO:0007386; P:compartment specification; IMP.
 CC GO: GO:0045944; P:positive regulation of transcription from P. .; IDA.
 CC InterPro: IPR002110; ANK.
 CC InterPro: IPR000152; Asx hydroxyl_S.
 CC InterPro: IPR000742; EGF_2.
 CC InterPro: IPR001881; EGF_Ca.
 CC InterPro: IPR001438; EGF_II.
 CC InterPro: IPR006209; EGF_III.
 CC InterPro: IPR002049; Laminin_EGF.
 CC InterPro: IPR008297; Notch.
 CC InterPro: IPR008800; Notch_dom.
 CC Pfam: PF00023; ank; 7.
 CC Pfam: PF00066; notch; 3.
 CC PIRSF: PIRSF002279; Notch; 1.
 CC PRINTS: PRO0010; EGFLOOD.
 CC PRINTS: PRO0011; EGFAMININ.
 CC PRINTS: PRO1452; NOTCH.
 CC SMART: SM00248; ANK; 6.
 CC SMART: SM00179; EGF_CA; 24.
 CC SMART: SM00004; NL; 2.
 CC PROSITE: PS50297; ANK_REPEAT; 1.
 CC PROSITE: PS50088; ANK_REPEAT; 2.
 CC PROSITE: PS00010; ASX_HYDROXYL; 22.
 CC PROSITE: PS00022; EGF_1; 34.
 CC PROSITE: PS01186; EGF_2; 27.
 CC PROSITE: PS50026; EGF_3; 36.
 CC PROSITE: PS01187; EGF_CA; 21.
 CC Receptor; Transcription regulation; Activator; Differentiation;
 CC Developmental protein; Repeat; ANK repeat; EGF-like domain;
 CC Transmembrane; Glycoprotein; Signal; Phosphorylation;
 CC Alternative splicing.
 CC SIGNAL 1 18 POTENTIAL.
 CC CHAIN 19 2531 NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 1.
 CC FT 1711 2531 NOTCH EXTRACELLULAR TRUNCATION.
 CC CHAIN 1744 2531 NOTCH INTRACELLULAR DOMAIN.

FT DOMAIN 19 1725 EXTRACELLULAR (POTENTIAL).
 Query Match 61.2%; Score 85; DB 1; Length 2531;
 Beet Local Similarity 45.2%; Pred. No. 5.9e-05;
 Matches 14; Conservative 2; Mismatches 15; Indels 0; Gaps 0;
 QY 1 CDXXXXXXKXGKXCDXXCNNAACXXDXDC 31
 DB 1449 CELPEQVDANKKVCNLCNNHACGWDGDC 1479
 RESULT 6
 NC1 RAT
 ID NC1 RAT STANDARD; PRT; 2531 AA.
 AC 007008;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Neurogenic locus notch homolog protein 1 precursor (Notch 1).
 GN NOTCH1.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OC NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Schwann cell;
 RX MEDLINE=92111383; PubMed=1764995;
 RA Weinmaster G., Roberts V.J., Lemke G.;
 RT "A homolog of Drosophila Notch expressed during mammalian
 RT development.";
 RL Development 113:199-205 (1991).
 RN [2]
 RP REVISIONS TO 1652-1653.
 RA Weinmaster G.;
 RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP FUNCTION.
 RX MEDLINE=21094508; PubMed=11182080;
 RA Tanigaki K., Nogaki F., Takahashi J., Tashiro K., Kurooka H.,
 RA Horijo T.;
 RT "Notch and Notch3 instructively restrict bFGF-responsive multipotent
 RT neural progenitor cells to an astroglial fate.";
 RL Neuron 29:45-55 (2001).
 RN [4]
 RP TISSUE SPECIFICITY.
 RX MEDLINE=93202015; PubMed=1295745;
 RA Weinmaster G., Roberts V.J., Lemke G.;
 RT "Notch2: a second mammalian Notch gene.";
 RL Development 116:931-941 (1992).
 RN [5]
 RP TISSUE SPECIFICITY.
 RX MEDLINE=21331789; PubMed=11438922;
 RA Irvan D.K., Zurcher S.D., Nguyen T., Weinmaster G., Kornblum H.I.;
 RT "Expression patterns of Notch1, Notch2, and Notch3 suggest multiple
 RT functional roles for the Notch-Delta signaling system during brain
 RT development.";
 RL J. Comp. Neurol. 436:167-181 (2001).
 CC -1- FUNCTION: Functions as a receptor for membrane-bound ligands
 CC tagged, tagged and Delta1 to regulate cell-fate determination.
 CC Upon ligand activation through the released notch intracellular
 CC domain (NICD) it forms a transcriptional activator complex with
 CC RBP-J kappa and activates genes of the enhancer of split locus.
 CC Affects the implementation of differentiation, proliferation and
 CC apoptotic programs (By similarity). Acts instructively to control
 CC the cell fate determination of CNS multipotent progenitor cells,
 CC resulting in astroglial induction and neuron/oligodendrocyte
 CC suppression.
 CC -1- SUBUNIT: Heterodimer of a C-terminal fragment N(TM) and a N-
 CC terminal fragment N(EC) which are probably linked by disulfide
 CC bonds (By similarity).
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Following
 CC proteolytical processing NICD is translocated to the nucleus (By

CC similarity).

CC -1- TISSUE SPECIFICITY: Expressed in the brain, kidney and spleen.

CC Expressed in postnatal central nervous system (CNS) germinal zones

CC and, in early postnatal life, within numerous cells throughout the

CC CNS. Found in both subventricular and ventricular germinal zones.

CC -1- DEVELOPMENTAL STAGE: In the embryo, highest levels occur between

CC days 12 and 14 and decrease rapidly to much lower levels in the

CC adult.

CC -1- PTM: Synthesized in the endoplasmic reticulum as an inactive form

CC which is proteolytically cleaved by a furin-like convertase for the

CC trans-Golgi network before it reaches the plasma membrane to yield

CC an active, ligand-accessible form. Cleavage results in a C-

CC terminal fragment N(TM) and a N-terminal fragment N(EC). Following

CC ligand binding, it is cleaved by TNF-alpha converting enzyme

CC (TACE) to yield a membrane-associated intermediate fragment called

CC notch extracellular truncation (NEXT). This fragment is then

CC cleaved by presenilin dependent gamma-secretase to release a

CC notch-derived peptide containing the intracellular domain (NICD)

CC from the membrane (By similarity).

CC -1- PTM: Phosphorylated (By similarity).

CC -1- SIMILARITY: Belongs to the NOTCH family.

CC -1- SIMILARITY: Contains 36 EGF-like domains.

CC -1- SIMILARITY: Contains 3 Lin/Notch repeats.

CC -1- SIMILARITY: Contains 5 ANK repeats.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as their content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC				This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - CC
CC				the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch). CC
CC				-----
CC	EMBL;	X57405;	CAA40667.1;	--.
DR	HSSP;	P00740;	1EDM.	
DR	InterPro;	IPR002110;	ANK.	
DR	InterPro;	IPR000152;	Aex_Hydroxyl_S.	
DR	InterPro;	IPR000742;	EGF 2.	
DR	InterPro;	IPR001891;	EGF Ca.	
DR	InterPro;	IPR001438;	EGF II.	
DR	InterPro;	IPR006209;	EGF_like	
DR	InterPro;	IPR002049;	Laminin_EGF.	
DR	InterPro;	IPR008297;	Notch.	
DR	PIfam;	PF000023;	ank; 6.	
DR	PIfam;	PR00008;	EGF; 35.	
DR	PIfam;	PF00066;	Notch; 3.	
DR	PIRSF;	PIRSF002279;	Notch; 1.	
DR	PRINTS;	PRO0010;	EGFBLOOD.	
DR	PRINTS;	PRO0011;	EGFLAMININ.	
DR	PRINTS;	PRO1452;	NOTCH.	
DR	SMART;	SMO0248;	ANK; 6.	
DR	SMART;	SMO0179;	EGF_CA; 25.	
DR	SMART;	SMO0004;	NL; 2.	
DR	PROSITE;	PS50297;	ANK REP REGION; 1.	
DR	PROSITE;	PS50088;	ANK REPEAT; 4.	
DR	PROSITE;	PS00010;	ASX_HYDROXYL; 22.	
DR	PROSITE;	PS00022;	EGF_1; 35.	
DR	PROSITE;	PS01186;	EGF_2; 26.	
DR	PROSITE;	PS50026;	EGF_3; 36.	
DR	PROSITE;	PS01187;	EGF_CA; 21.	
KW	Receptor;	Transcription regulation:	Activator; Differentiation;	
KW	Developmental protein:	Repeat:	ANK repeat; EGF-like domain;	
KW	Transmembrane;	Glycoprotein;	Signal; Phosphorylation.	
FT	SIGNAL	1		
FT		18		
FT	CHAIN	19	2531	
FT	CHAIN	1711	2531	
FT				
FT	CHAIN	1744	2531	
FT				
FT	DOMAIN	19	1723	
FT	TRANSMEM	1724	1746	
FT		1747	2531	
FT				
FT	DOMAIN			CYTOSOLASMIC (POTENTIAL).

FT	DOMAIN	20	58	BGF-LIKE 1.
FT	DOMAIN	59	99	BGF-LIKE 2.
FT	DOMAIN	102	139	BGF-LIKE 3.
FT	DOMAIN	140	136	BGF-LIKE 4.
FT	DOMAIN	178	216	BGF-LIKE 5.
FT	DOMAIN	218	255	BGF-LIKE 6.
FT	DOMAIN	257	293	BGF-LIKE 7.
FT	DOMAIN	295	333	BGF-LIKE 8.
FT	DOMAIN	335	371	BGF-LIKE 9.
FT	DOMAIN	372	410	BGF-LIKE 10.
FT	DOMAIN	412	450	BGF-LIKE 11.
FT	DOMAIN	452	488	BGF-LIKE 12.
FT	DOMAIN	490	526	BGF-LIKE 13.
FT	DOMAIN	528	564	BGF-LIKE 14.
FT	DOMAIN	566	601	BGF-LIKE 15.
FT	DOMAIN	603	639	BGF-LIKE 16.
FT	DOMAIN	641	676	BGF-LIKE 17.
FT	DOMAIN	678	714	BGF-LIKE 18.
FT	DOMAIN	716	751	BGF-LIKE 19.
FT	DOMAIN	753	789	BGF-LIKE 20.
FT	DOMAIN	791	827	BGF-LIKE 21.
FT	DOMAIN	829	867	BGF-LIKE 22.
FT	DOMAIN	869	905	BGF-LIKE 23.
FT	DOMAIN	907	943	BGF-LIKE 24.
FT	DOMAIN	985	981	BGF-LIKE 25.
FT	DOMAIN	983	1019	BGF-LIKE 26.
FT	DOMAIN	1021	1057	BGF-LIKE 27.
FT	DOMAIN	1059	1095	BGF-LIKE 28.
FT	DOMAIN	1097	1143	BGF-LIKE 29.
FT	DOMAIN	1145	1181	BGF-LIKE 30.
FT	DOMAIN	1183	1219	BGF-LIKE 31.
FT	DOMAIN	1221	1265	BGF-LIKE 32.
FT	DOMAIN	1267	1305	BGF-LIKE 33.
FT	DOMAIN	1307	1346	BGF-LIKE 34.
FT	DOMAIN	1348	1384	BGF-LIKE 35.
FT	DOMAIN	1387	1426	BGF-LIKE 36.
FT	REPEAT	1445	1480	LIN/NOTCH 1.
FT	REPEAT	1481	1522	LIN/NOTCH 2.
FT	REPEAT	1523	1562	LIN/NOTCH 3.
FT	REPEAT	1597	1646	ANK 1.
FT	REPEAT	1650	1680	ANK 2.
FT	REPEAT	1684	1713	ANK 3.
FT	REPEAT	1717	1746	ANK 4.
FT	REPEAT	1749	1778	ANK 5.
FT	DOMAIN	1730	1763	POLY-ALA.
FT	DOMAIN	1891	1894	POLY-GUT.
FT	DOMAIN	2258	2261	POLY-PRO.
FT	DOMAIN	2497	2500	CLEAV-SEB.
FT	SITE	1654	1655	CLEAV-SEB (BY SIMILARITY).
FT	DISULFID	24	37	BY SIMILARITY.
FT	DISULFID	31	46	BY SIMILARITY.
FT	DISULFID	48	57	BY SIMILARITY.
FT	DISULFID	63	74	BY SIMILARITY.
FT	DISULFID	68	87	BY SIMILARITY.
FT	DISULFID	89	98	BY SIMILARITY.
FT	DISULFID	106	117	BY SIMILARITY.
FT	DISULFID	111	127	BY SIMILARITY.
FT	DISULFID	129	138	BY SIMILARITY.
FT	DISULFID	144	155	BY SIMILARITY.
FT	DISULFID	149	164	BY SIMILARITY.
FT	DISULFID	166	175	BY SIMILARITY.
FT	DISULFID	182	195	BY SIMILARITY.
FT	DISULFID	189	204	BY SIMILARITY.
FT	DISULFID	206	215	BY SIMILARITY.
FT	DISULFID	222	233	BY SIMILARITY.
FT	DISULFID	227	243	BY SIMILARITY.

Query Match	61.2%	Score 85;	DB 1;	Length 2511;
Best Local Similarity	45.2%	Pred. No. 5.9e-05;		
Matches	14;	Conservative	2;	Mismatches 15;
				Indels 0;
				Gaps 0.
Y	1	CDXXXCXXKXGXGCDXXCNNAACXXDXGDC	31	

Db 1449 CELPECCEDAGNRYCNLCNNHACGWDGDC 1479

RESULT 7

NTC2_HUMAN STANDARD; PRT; 2471 AA.

ID NTC2_HUMAN 004721; Q99734; Q9H240;

AC 28-FEB-2003 (Rel. 41, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Neogenic locus notch homolog protein 2 precursor (Notch 2) (hN2).

GN NOTCH2.

OS Homo sapiens (human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

OX NCBI_TaxID=9606;

RN (1)

RP SEQUENCE FROM N.A.

RC TISSUE=Brain;

RA Blaumweller C.M., Mann R.S.;

RT "Complete human notch 2 (hN2) cDNA sequence";

RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.

RN (2)

RP SEQUENCE FROM N.A.

RC TISSUE=Breast tumor;

RA Correa R.G., Camargo A.A., Moreira E.S., Simpson A.J.G.;

RT "Human Notch2, a novel member of cell-fate determining NOTCH family";

RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.

RN (3)

RP SEQUENCE OF 967-1229 FROM N.A.

RC TISSUE=T-cell;

RA Lemasson I., Devaux C., Mesnard J.M.;

RT "Partial sequence of EGF-like repeat domain of human Notch2 mRNA.";

RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.

RN (4)

RP SEQUENCE OF 1810-2447 FROM N.A.

RC TISSUE=Brain;

RA Stifani S., Blaumweller C.M., Redhead N.J., Hill R.E.;

RT "Human homologue of a Drosophila enhancer of split gene product define a novel family of nuclear proteins.";

RL Nat. Genet. 2:119-127(1992).

RN (5)

RP POST-TRANSLATIONAL PROCESSING.

RA Blaumweller C.M., Qi H., Zagouras P., Artavanis-Tsakonas S.;

RT "Intracellular cleavage of Notch leads to a heterodimeric receptor on the plasma membrane.";

RL Cell 90:281-291(1997).

RN (6)

RP IDENTIFICATION OF LIGANDS.

RA Gray G.B., Mann R.S., Maltoidis E., Henrique D., Carcangiu M.-L., Banks A., Leiman J., Ward D., Ish-Horowitz D., Artavanis-Tsakonas S.;

RT "Human ligands of the Notch receptor.";

RL Am. J. Pathol. 154:785-794(1999)

CC -1- FUNCTION: Functions as a receptor for membrane-bound ligands tagged, jagged2 and Delta to regulate cell-fate determination. Upon ligand activation through the released notch intracellular domain (NICD) it forms a transcriptional activator complex with RBP-J kappa and activates genes of the enhancer of split locus. Affects the implementation of differentiation, proliferation and apoptotic programs (By similarity).

CC -1- SUBUNIT: Heterodimer of a C-terminal fragment N(TM) and a N-terminal fragment N(IEC) which are probably linked by disulfide bonds (By similarity).

CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Following proteolytical processing NICD is translocated to the nucleus.

CC -1- TISSUE SPECIFICITY: Expressed in the brain, heart, kidney, lung, skeletal muscle and liver.

CC -1- PTM: Synthesized in the endoplasmic reticulum as an inactive form which is proteolytically cleaved by a furin-like convertase in the trans-golgi network before it reaches the plasma membrane to yield an active, ligand-accessible form. Cleavage results in a C-terminal fragment N(TM) and a N-terminal fragment N(IEC). Following ligand binding, it is cleaved by TNF-alpha converting enzyme (TACE) to yield a membrane-associated intermediate fragment called notch extracellular truncation (NEXT). This fragment is then cleaved by presenilin dependent gamma-secretase to release a notch-derived peptide containing the intracellular domain (NICD) from the membrane (By similarity).

CC -1- PTM: Phosphorylated (By similarity).

CC -1- SIMILARITY: Belongs to the NOTCH family.

CC -1- SIMILARITY: Contains 35 EGF-like domains.

CC -1- SIMILARITY: Contains 2 Lin/Notch repeats.

CC -1- SIMILARITY: Contains 6 Ank repeats.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL; AF153535; AAG37073.1; -.

DR EMBL; U77493; AB19224.1; -.

DR HSSP; P00740; 1EDM.

DR GENE; HGNC:7882; NOTCH2.

DR MIM; 600275; -.

DR InterPro; IPR002110; ANK.

DR InterPro; IPR000152; Asx hydroxyl_S.

DR InterPro; IPR000742; EGF_2.

DR InterPro; IPR001881; EGF_Ca.

DR InterPro; IPR001438; EGF_II.

DR InterPro; IPR006209; EGF_like.

DR InterPro; IPR002049; Laminin_EGF.

DR InterPro; IPR008297; Notch.

DR InterPro; IPR008000; Notch_dom.

DR Pfam; PF000023; ank; 6.

DR Pfam; PF00008; EGF; 35.

DR Pfam; PF00066; notch; 2.

DR PIRSF; PIRSF002279; Notch; 1.

DR PRINTS; PR00010; EGFBL00D.

DR PRINTS; PR00011; EGF_LAMININ.

DR PRINTS; PR01452; NOTCH.

DR SMART; SM00248; ANK; 6.

DR SMART; SM00179; EGF_CA; 23.

DR SMART; SM00004; NL; 2.

DR PROSITE; PS50297; ANK REP REGION; 1.

DR PROSITE; PS50089; ANK REPEAT; 4.

DR PROSITE; PS00010; ASX HYDROXYL; 22.

DR PROSITE; PS00022; EGF_1; 34.

DR PROSITE; PS01186; EGF_2; 29.

DR PROSITE; PS50026; EGF_3; 35.

DR PROSITE; PS01187; EGF_CA; 22.

KM Receptor; Transcription regulation; Activator; Differentiation; Developmental protein; Repeat; ANK repeat; EGF-like domain; Transmembrane; Glycoprotein; Signal; Phosphorylation.

KM SIGNAL 1 25

FT CHAIN 26 2471

FT CHAIN 1666 2471

FT CHAIN 1697 2471

FT DOMAIN 26 1677

FT TRANSMEM 1678 1698

FT DOMAIN 1699 2471

FT DOMAIN 26 63

FT DOMAIN 64 102

FT DOMAIN 105 143

FT DOMAIN 144 180

FT DOMAIN 144 180

FT EGF-LIKE 4.

DR HSP; P00743; 1CCF.
 DR InterPro; IPR002110; ANK.
 DR InterPro; IPR000152; Asx_hydroxyl_5.
 DR InterPro; IPR000742; EGF_2.
 DR InterPro; IPR001881; EGF_Ca.
 DR InterPro; IPR001438; EGF_Li.
 DR InterPro; IPR006209; EGF_like.
 DR InterPro; IPR002049; LamInin_EGF.
 DR InterPro; IPR008297; Notch.
 DR InterPro; IPR000800; Notch_dom.
 DR Pfam; PF00023; ank; 6.
 DR Pfam; PF00008; EGF; 35.
 DR Pfam; PF00066; notch; 2.
 DR PIRSF; PIRSF002279; Notch; 1.
 DR PRINTS; PR00010; EGFBL00D.
 DR PRINTS; PR00011; EGFLAMININ.
 DR SMART; SM00248; NOTCH.
 DR SMART; SM00179; EGF_CA; 24.
 DR SMART; SM00004; NL; 2.
 DR PROSITE; PS50297; ANK_REPEAT; 1.
 DR PROSITE; PS50088; ANK_REPEAT; 4.
 DR PROSITE; PS00010; ASX_HYDROXYL; 22.
 DR PROSITE; PS00022; EGF_1; 34.
 DR PROSITE; PS01186; EGF_2; 26.
 DR PROSITE; PS50026; EGF_3; 35.
 DR PROSITE; PS01187; EGF_CA; 22.
 DR Receptor; Transcription regulation; Activator; Differentiation;
 KM Developmental protein; Repeat; ANK repeat; EGF-like domain;
 KW Transmembrane; Glycoprotein; Signal; Phosphorylation.
 FT SIGNAL 1 25
 FT CHAIN 26 2471
 FT CHAIN 1666 2471
 FT CHAIN 1697 2471
 FT DOMAIN 26 1677
 FT TRANSMEM 1678 1698
 FT DOMAIN 1699 2471
 FT DOMAIN 26 63
 FT DOMAIN 64 102
 FT DOMAIN 105 143
 FT DOMAIN 144 180
 FT DOMAIN 182 219
 FT DOMAIN 221 258
 FT DOMAIN 260 296
 FT DOMAIN 298 336
 FT DOMAIN 338 374
 FT DOMAIN 375 413
 FT DOMAIN 415 454
 FT DOMAIN 456 492
 FT DOMAIN 494 530
 FT DOMAIN 532 568
 FT DOMAIN 570 605
 FT DOMAIN 607 643
 FT DOMAIN 645 680
 FT DOMAIN 682 718
 FT DOMAIN 720 755
 FT DOMAIN 757 793
 FT DOMAIN 795 831
 FT DOMAIN 833 871
 FT DOMAIN 873 909
 FT DOMAIN 911 947
 FT DOMAIN 949 985
 FT DOMAIN 987 1023
 FT DOMAIN 1025 1061
 FT DOMAIN 1063 1099
 FT DOMAIN 1101 1147
 FT DOMAIN 1149 1185
 FT DOMAIN 1187 1223
 FT DOMAIN 1225 1262
 FT DOMAIN 1264 1302
 FT DOMAIN 1304 1343

FT DOMAIN 1374 1412 EGF-LIKE 35.
 FT DOMAIN 1645 1648 POLY-ALA.
 FT DOMAIN 1694 1697 POLY-LEU.
 FT DOMAIN 2426 2429 POLY-SER.
 FT DOMAIN 2446 2451 POLY-GLY.
 FT REPEAT 1420 1456 LIN/NOTCH 1.
 FT REPEAT 1503 1535 LIN/NOTCH 2.
 FT REPEAT 1827 1871 ANK 1.
 FT REPEAT 1876 1905 ANK 2.
 FT REPEAT 1909 1939 ANK 3.
 FT REPEAT 1943 1972 ANK 4.
 FT REPEAT 1976 2005 ANK 5.
 FT REPEAT 2009 2038 ANK 6.
 FT DISULFID 28 41 BY SIMILARITY.
 FT DISULFID 35 51 BY SIMILARITY.
 FT DISULFID 53 62 BY SIMILARITY.
 FT DISULFID 68 79 BY SIMILARITY.
 FT DISULFID 73 90 BY SIMILARITY.
 FT DISULFID 92 101 BY SIMILARITY.
 FT DISULFID 109 121 BY SIMILARITY.
 FT DISULFID 115 131 BY SIMILARITY.
 FT DISULFID 133 142 BY SIMILARITY.
 FT DISULFID 148 159 BY SIMILARITY.
 FT DISULFID 153 168 BY SIMILARITY.
 FT DISULFID 170 179 BY SIMILARITY.
 FT DISULFID 186 198 BY SIMILARITY.
 FT DISULFID 192 207 BY SIMILARITY.
 FT DISULFID 209 218 BY SIMILARITY.
 FT DISULFID 225 236 BY SIMILARITY.
 FT DISULFID 230 246 BY SIMILARITY.
 FT DISULFID 248 257 BY SIMILARITY.
 FT DISULFID 264 275 BY SIMILARITY.
 FT DISULFID 269 284 BY SIMILARITY.
 FT DISULFID 286 295 BY SIMILARITY.
 FT DISULFID 302 315 BY SIMILARITY.
 FT DISULFID 309 324 BY SIMILARITY.
 FT DISULFID 326 335 BY SIMILARITY.
 FT DISULFID 342 353 BY SIMILARITY.
 FT DISULFID 347 362 BY SIMILARITY.
 FT DISULFID 364 373 BY SIMILARITY.
 FT DISULFID 379 390 BY SIMILARITY.
 FT DISULFID 384 401 BY SIMILARITY.
 FT DISULFID 403 412 BY SIMILARITY.
 FT DISULFID 419 433 BY SIMILARITY.
 FT DISULFID 427 442 BY SIMILARITY.
 FT DISULFID 444 453 BY SIMILARITY.
 FT DISULFID 460 471 BY SIMILARITY.
 FT DISULFID 465 480 BY SIMILARITY.
 FT DISULFID 482 491 BY SIMILARITY.
 FT DISULFID 498 509 BY SIMILARITY.
 FT DISULFID 503 518 BY SIMILARITY.
 FT DISULFID 520 529 BY SIMILARITY.
 FT DISULFID 536 547 BY SIMILARITY.

Query Match 60.4%; Score 84; Length 2471;
 Best Local Similarity 45.2%; Pred. No. 8.1e-05;
 Matches 14; Conservative 2; Mismatches 15; Indels 0; Gaps 0;

QY 1 CDXXXCKXXKXGNCXKXCNNAKXDXGDC 31
 DB 1425 CUSQYCADRARDGICDEACNSHACQWDGDC 1455

RESULT 9
 NTCL_HUMAN STANDARD; PRT; 2556 AA.
 AC P46531;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 28-FEB-2003 (Rel. 41, last sequence update)
 DT 10-OCT-2003 (Rel. 42, last annotation update)
 DE Neurogenic locus notch homolog protein 1 precursor (Notch 1) (hml)
 DE (Translocation-associated notch protein TAN-1).
 GN NOTCH1 OR TAN1.

OS Homo sapiens (Human).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 OK NCBI_TaxID=9606;
 [1]
 RP SEQUENCE FROM N.A.
 RC Tissue=Brain;
 RA Mann R.S., Blumheller C.M., Zagouras P.;
 RT "Complete human notch 1 (hN1) cDNA sequence."
 RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
 [2]
 RP SEQUENCE OF 1-2444 FROM N.A.
 RX MEDLINE=91347367; PubMed=133192;
 RA Ehlisen L.W., Bird J., West D.C., Soreng A.L., Reynolds T.C.,
 RT Smith S.D., Sklar J.;
 RT "TAN-1, the human homolog of the Drosophila notch gene, is broken by
 RT chromosomal translocations in T lymphoblastic neoplasms."
 RL Cell 66:649-661(1991).
 [3]
 RP IDENTIFICATION OF LIGANDS.
 RX MEDLINE=99180765; PubMed=10079256;
 RA Gray G.E., Mann R.S., Mitsiadis E., Henrique D., Carcangiu M.-L.,
 RA Banks A., Leiman J., Ward D., Ish-Horowitz D., Artavanis-Tsakonas S.;
 RT "Human ligands of the Notch receptor."
 RL Am. J. Pathol. 154:785-794(1999).
 [4]
 RP INTERACTION WITH DTX1.
 RX MEDLINE=98250176; PubMed=9590294;
 RA Matsuno K., Eastman D., Mitsiadis T., Quinn A.M., Carcangiu M.L.,
 RA Orendlich P., Kadesch T., Artavanis-Tsakonas S.;
 RT "Human deltex is a conserved regulator of Notch signalling."
 RL Nat. Genet. 19:74-78(1998).
 -1- FUNCTION: Functions as a receptor for membrane-bound ligands
 Jagged1, Jagged2 and Deltal to regulate cell-fate determination.
 Upon ligand activation through the released notch intracellular
 domain (NICD) it forms a transcriptional activator complex with
 RBP-J kappa and activates genes of the enhancer of split locus.
 Affects the implementation of differentiation, proliferation and
 apoptotic programs. May be important for normal lymphocyte
 function. In altered form, may contribute to transformation or
 progression in some T-cell neoplasms. Involved in the maturation
 of both CD4+ and CD8+ cells in the thymus (By similarity).
 -1- SUBUNIT: Heterodimer of a C-terminal fragment N(FW) and a N-
 terminal fragment N(EC) which are probably linked by disulfide
 bonds (By similarity). Interacts with DTX1 and DTX2.
 -1- SUBCELLULAR LOCATION: Type I membrane protein. Following
 proteolytic processing NICD is translocated to the nucleus (By
 similarity).
 -1- TISSUE SPECIFICITY: In fetal tissues most abundant in spleen,
 brain stem and lung. Also present in most adult tissues where it
 is found mainly in lymphoid tissues.
 -1- PTM: Synthesized in the endoplasmic reticulum as an inactive form
 which is proteolytically cleaved by a furin-like convertase in the
 trans-Golgi network before it reaches the plasma membrane to yield
 an active, ligand-accessible form. Cleavage results in a C-
 terminal fragment N(FW) and a N-terminal fragment N(EC). Following
 ligand binding, it is cleaved by TNF-alpha converting enzyme
 (TACE) to yield a membrane-associated intermediate fragment called
 notch extracellular truncation (NEXT). This fragment is then
 cleaved by presenilin dependent gamma-secretase to release a
 notch-derived peptide containing the intracellular domain (NICD)
 from the membrane (By similarity).
 -1- PTM: Phosphorylated (By similarity).
 -1- DISEASE: NOTCH1 truncation is associated with T-cell acute
 lymphoblastic leukemia.
 -1- SIMILARITY: Belongs to the NOTCH family.
 -1- SIMILARITY: Contains 36 EGF-like domains.
 -1- SIMILARITY: Contains 3 Lin/Notch repeats.
 -1- SIMILARITY: Contains 5 ANK repeats.
 CC This Swiss-Prot entry is copyrighted. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AF308602; AAC33848.1; -;
 CC EMBL; M73980; AAA60614.1; -;
 CC HSSP; P00740; 1EDM.
 CC Genew: HGNC:7881; NOTCH1.
 CC MIM; 190198; -;
 CC GO; GO:0016021; C:integral to membrane; NAS.
 CC GO; GO:0003793; F:defense/immunity protein activity; NAS.
 CC GO; GO:0006955; P:immune response; NAS.
 CC InterPro; IPR002110; ANK.
 CC InterPro; IPR000152; Asx_hydroxyl_S.
 CC InterPro; IPR000742; EGF_2.
 CC InterPro; IPR001881; EGF_CA.
 CC InterPro; IPR001438; EGF_II.
 CC InterPro; IPR006209; EGF_like.
 CC InterPro; IPR002049; Laminin_EGF.
 CC InterPro; IPR008297; Notch.
 CC InterPro; IPR000800; Notch_dom.
 CC Pfam; PF00023; ank; 6.
 CC Pfam; PF00008; EGF; 35.
 CC Pfam; PF00066; notch; 3.
 CC PIRSF; PIRSF002279; Notch; 1.
 CC PRINTS; PR00010; EGFBLD.
 CC PRINTS; PR00011; EGF_LAMININ.
 CC PRINTS; PR01452; NOTCH.
 CC SMART; SM00248; ANK; 6.
 CC SMART; SM00179; EGF_CA; 23.
 CC SMART; SM00004; NU; 3.
 CC PROSITE; PSS0297; ANK_REPEAT_1.
 CC PROSITE; PSS0088; ANK_REPEAT_4.
 CC PROSITE; PS00010; ASX_HYDROXYL; 20.
 CC PROSITE; PS00022; EGF_1; 34.
 CC PROSITE; PS01186; EGF_2; 26.
 CC PROSITE; PSS0026; EGF_3; 36.
 CC PROSITE; PS01187; EGF_CA; 18.
 CC Receptor; Transcription regulation; Activator; Differentiation;
 CC Developmental protein; Repeat; ANK repeat; EGF-like domain;
 CC Transmembrane; Glycoprotein; Signal; Phosphorylation.
 CC FT CHAIN 1 18
 CC FT CHAIN 19 2556
 CC FT CHAIN 1722 2556
 CC FT CHAIN 1755 2556
 CC FT DOMAIN 19 1736
 CC FT TRANSMEM 1737 1757
 CC FT DOMAIN 1758 2556
 CC FT DOMAIN 20 58
 CC FT DOMAIN 59 99
 CC FT DOMAIN 102 139
 CC FT DOMAIN 140 176
 CC FT DOMAIN 178 216
 CC FT DOMAIN 218 255
 CC FT DOMAIN 257 293
 CC FT DOMAIN 295 333
 CC FT DOMAIN 335 371
 CC FT DOMAIN 372 410
 CC FT DOMAIN 412 450
 CC FT DOMAIN 452 488
 CC FT DOMAIN 490 526
 CC FT DOMAIN 528 564
 CC FT DOMAIN 566 601
 CC FT DOMAIN 603 639
 CC FT DOMAIN 641 676
 CC FT DOMAIN 678 714
 CC FT DOMAIN 716 751
 CC FT DOMAIN 753 789
 CC FT DOMAIN 791 827
 CC FT DOMAIN 829 868
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AF308602; AAC33848.1; -;
 CC EMBL; M73980; AAA60614.1; -;
 CC HSSP; P00740; 1EDM.
 CC Genew: HGNC:7881; NOTCH1.
 CC MIM; 190198; -;
 CC GO; GO:0016021; C:integral to membrane; NAS.
 CC GO; GO:0003793; F:defense/immunity protein activity; NAS.
 CC GO; GO:0006955; P:immune response; NAS.
 CC InterPro; IPR002110; ANK.
 CC InterPro; IPR000152; Asx_hydroxyl_S.
 CC InterPro; IPR000742; EGF_2.
 CC InterPro; IPR001881; EGF_CA.
 CC InterPro; IPR001438; EGF_II.
 CC InterPro; IPR006209; EGF_like.
 CC InterPro; IPR002049; Laminin_EGF.
 CC InterPro; IPR008297; Notch.
 CC InterPro; IPR000800; Notch_dom.
 CC Pfam; PF00023; ank; 6.
 CC Pfam; PF00008; EGF; 35.
 CC Pfam; PF00066; notch; 3.
 CC PIRSF; PIRSF002279; Notch; 1.
 CC PRINTS; PR00010; EGFBLD.
 CC PRINTS; PR00011; EGF_LAMININ.
 CC PRINTS; PR01452; NOTCH.
 CC SMART; SM00248; ANK; 6.
 CC SMART; SM00179; EGF_CA; 23.
 CC SMART; SM00004; NU; 3.
 CC PROSITE; PSS0297; ANK_REPEAT_1.
 CC PROSITE; PSS0088; ANK_REPEAT_4.
 CC PROSITE; PS00010; ASX_HYDROXYL; 20.
 CC PROSITE; PS00022; EGF_1; 34.
 CC PROSITE; PS01186; EGF_2; 26.
 CC PROSITE; PSS0026; EGF_3; 36.
 CC PROSITE; PS01187; EGF_CA; 18.
 CC Receptor; Transcription regulation; Activator; Differentiation;
 CC Developmental protein; Repeat; ANK repeat; EGF-like domain;
 CC Transmembrane; Glycoprotein; Signal; Phosphorylation.
 CC FT CHAIN 1 18
 CC FT CHAIN 19 2556
 CC FT CHAIN 1722 2556
 CC FT CHAIN 1755 2556
 CC FT DOMAIN 19 1736
 CC FT TRANSMEM 1737 1757
 CC FT DOMAIN 1758 2556
 CC FT DOMAIN 20 58
 CC FT DOMAIN 59 99
 CC FT DOMAIN 102 139
 CC FT DOMAIN 140 176
 CC FT DOMAIN 178 216
 CC FT DOMAIN 218 255
 CC FT DOMAIN 257 293
 CC FT DOMAIN 295 333
 CC FT DOMAIN 335 371
 CC FT DOMAIN 372 410
 CC FT DOMAIN 412 450
 CC FT DOMAIN 452 488
 CC FT DOMAIN 490 526
 CC FT DOMAIN 528 564
 CC FT DOMAIN 566 601
 CC FT DOMAIN 603 639
 CC FT DOMAIN 641 676
 CC FT DOMAIN 678 714
 CC FT DOMAIN 716 751
 CC FT DOMAIN 753 789
 CC FT DOMAIN 791 827
 CC FT DOMAIN 829 868
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AF308602; AAC33848.1; -;
 CC EMBL; M73980; AAA60614.1; -;
 CC HSSP; P00740; 1EDM.
 CC Genew: HGNC:7881; NOTCH1.
 CC MIM; 190198; -;
 CC GO; GO:0016021; C:integral to membrane; NAS.
 CC GO; GO:0003793; F:defense/immunity protein activity; NAS.
 CC GO; GO:0006955; P:immune response; NAS.
 CC InterPro; IPR002110; ANK.
 CC InterPro; IPR000152; Asx_hydroxyl_S.
 CC InterPro; IPR000742; EGF_2.
 CC InterPro; IPR001881; EGF_CA.
 CC InterPro; IPR001438; EGF_II.
 CC InterPro; IPR006209; EGF_like.
 CC InterPro; IPR002049; Laminin_EGF.
 CC InterPro; IPR008297; Notch.
 CC InterPro; IPR000800; Notch_dom.
 CC Pfam; PF00023; ank; 6.
 CC Pfam; PF00008; EGF; 35.
 CC Pfam; PF00066; notch; 3.
 CC PIRSF; PIRSF002279; Notch; 1.
 CC PRINTS; PR00010; EGFBLD.
 CC PRINTS; PR00011; EGF_LAMININ.
 CC PRINTS; PR01452; NOTCH.
 CC SMART; SM00248; ANK; 6.
 CC SMART; SM00179; EGF_CA; 23.
 CC SMART; SM00004; NU; 3.
 CC PROSITE; PSS0297; ANK_REPEAT_1.
 CC PROSITE; PSS0088; ANK_REPEAT_4.
 CC PROSITE; PS00010; ASX_HYDROXYL; 20.
 CC PROSITE; PS00022; EGF_1; 34.
 CC PROSITE; PS01186; EGF_2; 26.
 CC PROSITE; PSS0026; EGF_3; 36.
 CC PROSITE; PS01187; EGF_CA; 18.
 CC Receptor; Transcription regulation; Activator; Differentiation;
 CC Developmental protein; Repeat; ANK repeat; EGF-like domain;
 CC Transmembrane; Glycoprotein; Signal; Phosphorylation.
 CC FT CHAIN 1 18
 CC FT CHAIN 19 2556
 CC FT CHAIN 1722 2556
 CC FT CHAIN 1755 2556
 CC FT DOMAIN 19 1736
 CC FT TRANSMEM 1737 1757
 CC FT DOMAIN 1758 2556
 CC FT DOMAIN 20 58
 CC FT DOMAIN 59 99
 CC FT DOMAIN 102 139
 CC FT DOMAIN 140 176
 CC FT DOMAIN 178 216
 CC FT DOMAIN 218 255
 CC FT DOMAIN 257 293
 CC FT DOMAIN 295 333
 CC FT DOMAIN 335 371
 CC FT DOMAIN 372 410
 CC FT DOMAIN 412 450
 CC FT DOMAIN 452 488
 CC FT DOMAIN 490 526
 CC FT DOMAIN 528 564
 CC FT DOMAIN 566 601
 CC FT DOMAIN 603 639
 CC FT DOMAIN 641 676
 CC FT DOMAIN 678 714
 CC FT DOMAIN 716 751
 CC FT DOMAIN 753 789
 CC FT DOMAIN 791 827
 CC FT DOMAIN 829 868
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AF308602; AAC33848.1; -;
 CC EMBL; M73980; AAA60614.1; -;
 CC HSSP; P00740; 1EDM.
 CC Genew: HGNC:7881; NOTCH1.
 CC MIM; 190198; -;
 CC GO; GO:0016021; C:integral to membrane; NAS.
 CC GO; GO:0003793; F:defense/immunity protein activity; NAS.
 CC GO; GO:0006955; P:immune response; NAS.
 CC InterPro; IPR002110; ANK.
 CC InterPro; IPR000152; Asx_hydroxyl_S.
 CC InterPro; IPR000742; EGF_2.
 CC InterPro; IPR001881; EGF_CA.
 CC InterPro; IPR001438; EGF_II.
 CC InterPro; IPR006209; EGF_like.
 CC InterPro; IPR002049; Laminin_EGF.
 CC InterPro; IPR008297; Notch.
 CC InterPro; IPR000800; Notch_dom.
 CC Pfam; PF00023; ank; 6.
 CC Pfam; PF00008; EGF; 35.
 CC Pfam; PF00066; notch; 3.
 CC PIRSF; PIRSF002279; Notch; 1.
 CC PRINTS; PR00010; EGFBLD.
 CC PRINTS; PR00011; EGF_LAMININ.
 CC PRINTS; PR01452; NOTCH.
 CC SMART; SM00248; ANK; 6.
 CC SMART; SM00179; EGF_CA; 23.
 CC SMART; SM00004; NU; 3.
 CC PROSITE; PSS0297; ANK_REPEAT_1.
 CC PROSITE; PSS0088; ANK_REPEAT_4.
 CC PROSITE; PS00010; ASX_HYDROXYL; 20.
 CC PROSITE; PS00022; EGF_1; 34.
 CC PROSITE; PS01186; EGF_2; 26.
 CC PROSITE; PSS0026; EGF_3; 36.
 CC PROSITE; PS01187; EGF_CA; 18.
 CC Receptor; Transcription regulation; Activator; Differentiation;
 CC Developmental protein; Repeat; ANK repeat; EGF-like domain;
 CC Transmembrane; Glycoprotein; Signal; Phosphorylation.
 CC FT CHAIN 1 18
 CC FT CHAIN 19 2556
 CC FT CHAIN 1722 2556
 CC FT CHAIN 1755 2556
 CC FT DOMAIN 19 1736
 CC FT TRANSMEM 1737 1757
 CC FT DOMAIN 1758 2556
 CC FT DOMAIN 20 58
 CC FT DOMAIN 59 99
 CC FT DOMAIN 102 139
 CC FT DOMAIN 140 176
 CC FT DOMAIN 178 216
 CC FT DOMAIN 218 255
 CC FT DOMAIN 257 293
 CC FT DOMAIN 295 333
 CC FT DOMAIN 335 371
 CC FT DOMAIN 372 410
 CC FT DOMAIN 412 450
 CC FT DOMAIN 452 488
 CC FT DOMAIN 490 526
 CC FT DOMAIN 528 564
 CC FT DOMAIN 566 601
 CC FT DOMAIN 603 639
 CC FT DOMAIN 641 676
 CC FT DOMAIN 678 714
 CC FT DOMAIN 716 751
 CC FT DOMAIN 753 789
 CC FT DOMAIN 791 827
 CC FT DOMAIN 829 868
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AF308602; AAC33848.1; -;
 CC EMBL; M73980; AAA60614.1; -;
 CC HSSP; P00740; 1EDM.
 CC Genew: HGNC:7881; NOTCH1.
 CC MIM; 190198; -;
 CC GO; GO:0016021; C:integral to membrane; NAS.
 CC GO; GO:0003793; F:defense/immunity protein activity; NAS.
 CC GO; GO:0006955; P:immune response; NAS.
 CC InterPro; IPR002110; ANK.
 CC InterPro; IPR000152; Asx_hydroxyl_S.
 CC InterPro; IPR000742; EGF_2.
 CC InterPro; IPR001881; EGF_CA.
 CC InterPro; IPR001438; EGF_II.
 CC InterPro; IPR006209; EGF_like.
 CC InterPro; IPR002049; Laminin_EGF.
 CC InterPro; IPR008297; Notch.
 CC InterPro; IPR000800; Notch_dom.
 CC Pfam; PF00023; ank; 6.
 CC Pfam; PF00008; EGF; 35.
 CC Pfam; PF00066; notch; 3.
 CC PIRSF; PIRSF002279; Notch; 1.
 CC PRINTS; PR00010; EGFBLD.
 CC PRINTS; PR00011; EGF_LAMININ.
 CC PRINTS; PR01452; NOTCH.
 CC SMART; SM00248; ANK; 6.
 CC SMART; SM00179; EGF_CA; 23.
 CC SMART; SM00004; NU; 3.
 CC PROSITE; PSS0297; ANK_REPEAT_1.
 CC PROSITE; PSS0088; ANK_REPEAT_4.
 CC PROSITE; PS00010; ASX_HYDROXYL; 20.
 CC PROSITE; PS00022; EGF_1; 34.
 CC PROSITE; PS01186; EGF_2; 26.
 CC PROSITE; PSS0026; EGF_3; 36.
 CC PROSITE; PS01187; EGF_CA; 18.
 CC Receptor; Transcription regulation; Activator; Differentiation;
 CC Developmental protein; Repeat; ANK repeat; EGF-like domain;
 CC Transmembrane; Glycoprotein; Signal; Phosphorylation.
 CC FT CHAIN 1 18
 CC FT CHAIN 19 2556
 CC FT CHAIN 1722 2556
 CC FT CHAIN 1755 2556
 CC FT DOMAIN 19 1736
 CC FT TRANSMEM 1737 1757
 CC FT DOMAIN 1758 2556
 CC FT DOMAIN 20 58
 CC FT DOMAIN 59 99
 CC FT DOMAIN 102 139
 CC FT DOMAIN 140 176
 CC FT DOMAIN 178 216
 CC FT DOMAIN 218 255
 CC FT DOMAIN 257 293
 CC FT DOMAIN 295 333
 CC FT DOMAIN 335 371
 CC FT DOMAIN 372 410
 CC FT DOMAIN 412 450
 CC FT DOMAIN 452 488
 CC FT DOMAIN 490 526
 CC FT DOMAIN 528 564
 CC FT DOMAIN 566 601
 CC FT DOMAIN 603 639
 CC FT DOMAIN 641 676
 CC FT DOMAIN 678 714
 CC FT DOMAIN 716 751
 CC FT DOMAIN 753 789
 CC FT DOMAIN 791 827
 CC FT DOMAIN 829 868
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AF308602; AAC33848.1; -;
 CC EMBL; M73980; AAA60614.1; -;
 CC HSSP; P00740; 1EDM.
 CC Genew: HGNC:7881; NOTCH1.
 CC MIM; 190198; -;
 CC GO; GO:0016021; C:integral to membrane; NAS.
 CC GO; GO:0003793; F:defense/immunity protein activity; NAS.
 CC GO; GO:0006955; P:immune response; NAS.
 CC InterPro; IPR002110; ANK.
 CC InterPro; IPR000152; Asx_hydroxyl_S.
 CC InterPro; IPR000742; EGF_2.
 CC InterPro; IPR001881; EGF_CA.
 CC InterPro; IPR001438; EGF_II.
 CC InterPro; IPR006209; EGF_like.
 CC InterPro; IPR002049; Laminin_EGF.
 CC InterPro; IPR008297; Notch.
 CC InterPro; IPR000800; Notch_dom.
 CC Pfam; PF00023; ank; 6.
 CC Pfam; PF00008; EGF; 35.
 CC Pfam; PF00066; notch; 3.
 CC PIRSF; PIRSF002279; Notch; 1.
 CC PRINTS; PR00010; EGFBLD.
 CC PRINTS; PR00011; EGF_LAMININ.
 CC PRINTS; PR01452; NOTCH.
 CC SMART; SM00248; ANK; 6.
 CC SMART; SM00179; EGF_CA; 23.
 CC SMART; SM00004; NU; 3.
 CC PROSITE; PSS0297; ANK_REPEAT_1.
 CC PROSITE; PSS0088; ANK_REPEAT_4.
 CC PROSITE; PS00010; ASX_HYDROXYL; 20.
 CC PROSITE; PS00022; EGF_1; 34.
 CC PROSITE; PS01186; EGF_2; 26.
 CC PROSITE; PSS0026; EGF_3; 36.
 CC PROSITE; PS01187; EGF_CA; 18.
 CC Receptor; Transcription regulation; Activator; Differentiation;
 CC Developmental protein; Repeat; ANK repeat; EGF-like domain;
 CC Transmembrane; Glycoprotein; Signal; Phosphorylation.
 CC FT CHAIN 1 18
 CC FT CHAIN 19 2556
 CC FT CHAIN 1722 2556
 CC FT CHAIN 1755 2556
 CC FT DOMAIN 19 1736
 CC FT TRANSMEM 1737 1757
 CC FT DOMAIN 1758 2556
 CC FT DOMAIN 20 58
 CC FT DOMAIN 59 99
 CC FT DOMAIN 102 139
 CC FT DOMAIN 140 176
 CC FT DOMAIN 178 216
 CC FT DOMAIN 218 255
 CC FT DOMAIN 257 293
 CC FT DOMAIN 295 333
 CC FT DOMAIN 335 371
 CC FT DOMAIN 372 410
 CC FT DOMAIN 412 450
 CC FT DOMAIN 452 488
 CC FT DOMAIN 490 526
 CC FT DOMAIN 528 564
 CC FT DOMAIN 566 601
 CC FT DOMAIN 603 639
 CC FT DOMAIN 641 676
 CC FT DOMAIN 678 714
 CC FT DOMAIN 716 751
 CC FT DOMAIN 753 789
 CC FT DOMAIN 791 827
 CC FT DOMAIN 829 868
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AF308602; AAC33848.1; -;
 CC EMBL; M73980; AAA60614.1; -;
 CC HSSP; P00740; 1EDM.
 CC Genew: HGNC:7881; NOTCH1.
 CC MIM; 190198; -;
 CC GO; GO:0016021; C:integral to membrane; NAS.
 CC GO; GO:0003793; F:defense/immunity protein activity; NAS.
 CC GO; GO:0006955; P:immune response; NAS.
 CC InterPro; IPR002110; ANK.
 CC InterPro; IPR000152; Asx_hydroxyl_S.
 CC InterPro; IPR000742; EGF_2.
 CC InterPro; IPR001881; EGF_CA.
 CC InterPro; IPR001438; EGF_II.
 CC InterPro; IPR006209; EGF_like.
 CC InterPro; IPR002049; Laminin_EGF.
 CC InterPro; IPR008297; Notch.
 CC InterPro; IPR000800; Notch_dom.
 CC Pfam; PF00023; ank; 6.
 CC Pfam; PF00008; EGF; 35.
 CC Pfam; PF00066; notch; 3.
 CC PIRSF; PIRSF002279; Notch; 1.
 CC PRINTS; PR00010; EGFBLD.
 CC PRINTS; PR00011; EGF_LAMININ.
 CC PRINTS; PR01452; NOTCH.
 CC SMART; SM00248; ANK; 6.
 CC SMART; SM00179; EGF_CA; 23.
 CC SMART; SM00004; NU; 3.
 CC PROSITE; PSS0297; ANK_REPEAT_1.
 CC PROSITE; PSS0088; ANK_REPEAT_4.
 CC PROSITE; PS00010; ASX_HYDROXYL; 20.
 CC PROSITE; PS00022; EGF_1; 34.
 CC PROSITE; PS01186; EGF_2; 26.
 CC PROSITE; PSS0026; EGF_3; 36.
 CC PROSITE; PS01187; EGF_CA; 18.
 CC Receptor; Transcription regulation; Activator; Differentiation;
 CC Developmental protein; Repeat; ANK repeat; EGF-like domain;
 CC Transmembrane; Glycoprotein; Signal; Phosphorylation.
 CC FT CHAIN 1 18
 CC FT CHAIN 19 2556
 CC FT CHAIN 1722 2556
 CC FT CHAIN 1755 2556
 CC FT DOMAIN 19 1736
 CC FT TRANSMEM 1737 1757
 CC FT DOMAIN 1758 2556
 CC FT DOMAIN 20 58
 CC FT DOMAIN 59 99
 CC FT DOMAIN 102 139
 CC FT DOMAIN 140 176
 CC FT DOMAIN 178 216
 CC FT DOMAIN 218 255
 CC FT DOMAIN 257 293
 CC FT DOMAIN 295 333
 CC FT DOMAIN 335 371
 CC FT DOMAIN 372 410
 CC FT DOMAIN 412 450
 CC FT DOMAIN 452 488
 CC FT DOMAIN 490 526
 CC FT DOMAIN 528 564
 CC FT DOMAIN 566 601
 CC FT DOMAIN 603 639
 CC FT DOMAIN 641 676
 CC FT DOMAIN 678 714
 CC FT DOMAIN 716 751
 CC FT DOMAIN 753 789
 CC FT DOMAIN 791 827
 CC FT DOMAIN 829 868
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AF308602; AAC33848.1; -;
 CC EMBL; M73980; AAA60614.1; -;
 CC HSSP; P00740; 1EDM.
 CC Genew: HGNC:7881; NOTCH1.
 CC MIM; 190198; -;
 CC GO; GO:0016021; C:integral to membrane; NAS.
 CC GO; GO:0003793; F:defense/immunity protein activity; NAS.
 CC GO; GO:0006955; P:immune response; NAS.
 CC InterPro; IPR002110; ANK.
 CC InterPro; IPR000152; Asx_hydroxyl_S.
 CC InterPro; IPR000742; EGF_2.
 CC InterPro; IPR001881; EGF_CA.
 CC InterPro; IPR001438; EGF_II.
 CC InterPro; IPR006209; EGF_like.
 CC InterPro; IPR002049; Laminin_EGF.
 CC InterPro; IPR008297; Notch.
 CC InterPro; IPR000800; Notch_dom.
 CC Pfam; PF00023; ank; 6.
 CC Pfam; PF00008; EGF; 35.
 CC Pfam; PF00066; notch; 3.
 CC PIRSF; PIRSF002279; Notch; 1.
 CC PRINTS; PR00010; EGFBLD.
 CC PRINTS; PR00011; EGF_LAMININ.
 CC PRINTS; PR01452; NOTCH.
 CC SMART; SM00248; ANK; 6.
 CC SMART; SM00179; EGF_CA; 23.
 CC SMART; SM00004; NU; 3.
 CC PROSITE; PSS0297; ANK_REPEAT_1.
 CC PROSITE; PSS0088; ANK_REPEAT_4.
 CC PROSITE; PS00010; ASX_HYDROXYL; 20.
 CC PROSITE; PS00022; EGF_1; 34.
 CC PROSITE; PS01186; EGF_2; 26.
 CC PROSITE; PSS0026; EGF_3; 36.
 CC PROSITE; PS01187; EGF_CA; 18.
 CC Receptor; Transcription regulation; Activator; Differentiation;
 CC Developmental protein; Repeat; ANK repeat; EGF-like domain;
 CC Transmembrane; Glycoprotein; Signal; Phosphorylation.
 CC FT CHAIN 1 18
 CC FT CHAIN 19 2556
 CC FT CHAIN 1722 2556
 CC FT CHAIN 1755 2556
 CC FT DOMAIN 19 1736
 CC FT TRANSMEM 1737 1757
 CC FT DOMAIN 1758 2556
 CC FT DOMAIN 20 58
 CC FT DOMAIN 59 99
 CC FT DOMAIN 102 139
 CC FT DOMAIN 140 176
 CC FT DOMAIN 178 216
 CC FT DOMAIN 218 255
 CC FT DOMAIN 257 293
 CC FT DOMAIN 295 333
 CC FT DOMAIN 335 371
 CC FT DOMAIN 372 410

```

FT  DOMAIN  870  906  EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).
FT  DOMAIN  908  944  EGF-LIKE 24.
FT  DOMAIN  946  982  EGF-LIKE 25, CALCIUM-BINDING (POTENTIAL).
FT  DOMAIN  984  1020  EGF-LIKE 26.
FT  DOMAIN  1022  1058  EGF-LIKE 27.
FT  DOMAIN  1060  1096  EGF-LIKE 28.
FT  DOMAIN  1098  1144  EGF-LIKE 29.
FT  DOMAIN  1146  1182  EGF-LIKE 30.
FT  DOMAIN  1184  1220  EGF-LIKE 31, CALCIUM-BINDING (POTENTIAL).
FT  DOMAIN  1222  1266  EGF-LIKE 32, CALCIUM-BINDING (POTENTIAL).
FT  DOMAIN  1268  1306  EGF-LIKE 33.
FT  DOMAIN  1308  1347  EGF-LIKE 34.
FT  DOMAIN  1349  1385  EGF-LIKE 35.
FT  DOMAIN  1388  1427  EGF-LIKE 36.
FT  REPEAT  1446  1481  LIN/NOTCH 1.
FT  REPEAT  1482  1523  LIN/NOTCH 2.
FT  REPEAT  1524  1563  LIN/NOTCH 3.
FT  REPEAT  1563  1597  ANK 1.
FT  REPEAT  1591  1624  ANK 2.
FT  REPEAT  1624  1657  ANK 3.
FT  REPEAT  1657  1691  ANK 4.
FT  REPEAT  1691  1724  ANK 5.
FT  DOMAIN  1729  1762  POLY-VAL.
FT  DOMAIN  1762  1792  POLY-ARG.
FT  DOMAIN  1792  1824  POLY-PRO.
FT  DOMAIN  1824  1857  POLY-ALA.
FT  DOMAIN  1857  1890  POLY-GLU.
FT  DOMAIN  1890  1923  POLY-GLY.
FT  DOMAIN  1923  1956  POLY-GLN.
FT  DOMAIN  1956  1989  POLY-SER.
FT  SITE  1989  2022  CLEAVAGE (BY FURIN-LIKE PROTEASE) (BY
FT  SITE  2022  2055  SIMILARITY).
FT  DISULFID  2055  2088  BY SIMILARITY.
FT  DISULFID  2088  2121  BY SIMILARITY.
FT  DISULFID  2121  2154  BY SIMILARITY.
FT  DISULFID  2154  2187  BY SIMILARITY.
FT  DISULFID  2187  2220  BY SIMILARITY.
FT  DISULFID  2220  2253  BY SIMILARITY.
FT  DISULFID  2253  2286  BY SIMILARITY.
FT  DISULFID  2286  2319  BY SIMILARITY.
FT  DISULFID  2319  2352  BY SIMILARITY.
FT  DISULFID  2352  2385  BY SIMILARITY.
FT  DISULFID  2385  2418  BY SIMILARITY.
FT  DISULFID  2418  2451  BY SIMILARITY.
FT  DISULFID  2451  2484  BY SIMILARITY.
FT  DISULFID  2484  2517  BY SIMILARITY.
FT  DISULFID  2517  2550  BY SIMILARITY.
FT  DISULFID  2550  2583  BY SIMILARITY.
FT  DISULFID  2583  2616  BY SIMILARITY.
FT  DISULFID  2616  2649  BY SIMILARITY.
FT  DISULFID  2649  2682  BY SIMILARITY.
FT  DISULFID  2682  2715  BY SIMILARITY.
FT  DISULFID  2715  2748  BY SIMILARITY.
FT  DISULFID  2748  2781  BY SIMILARITY.
FT  DISULFID  2781  2814  BY SIMILARITY.
FT  DISULFID  2814  2847  BY SIMILARITY.
FT  DISULFID  2847  2880  BY SIMILARITY.
FT  DISULFID  2880  2913  BY SIMILARITY.
FT  DISULFID  2913  2946  BY SIMILARITY.
FT  DISULFID  2946  2979  BY SIMILARITY.
FT  DISULFID  2979  3012  BY SIMILARITY.
FT  DISULFID  3012  3045  BY SIMILARITY.
FT  DISULFID  3045  3078  BY SIMILARITY.
FT  DISULFID  3078  3111  BY SIMILARITY.
FT  DISULFID  3111  3144  BY SIMILARITY.
FT  DISULFID  3144  3177  BY SIMILARITY.
FT  DISULFID  3177  3210  BY SIMILARITY.
FT  DISULFID  3210  3243  BY SIMILARITY.
FT  DISULFID  3243  3276  BY SIMILARITY.
FT  DISULFID  3276  3309  BY SIMILARITY.
FT  DISULFID  3309  3342  BY SIMILARITY.
FT  DISULFID  3342  3375  BY SIMILARITY.
FT  DISULFID  3375  3408  BY SIMILARITY.
FT  DISULFID  3408  3441  BY SIMILARITY.
FT  DISULFID  3441  3474  BY SIMILARITY.
FT  DISULFID  3474  3507  BY SIMILARITY.
FT  DISULFID  3507  3540  BY SIMILARITY.
FT  DISULFID  3540  3573  BY SIMILARITY.
FT  DISULFID  3573  3606  BY SIMILARITY.
FT  DISULFID  3606  3639  BY SIMILARITY.
FT  DISULFID  3639  3672  BY SIMILARITY.
FT  DISULFID  3672  3705  BY SIMILARITY.
FT  DISULFID  3705  3738  BY SIMILARITY.
FT  DISULFID  3738  3771  BY SIMILARITY.
FT  DISULFID  3771  3804  BY SIMILARITY.
FT  DISULFID  3804  3837  BY SIMILARITY.
FT  DISULFID  3837  3870  BY SIMILARITY.
FT  DISULFID  3870  3903  BY SIMILARITY.
FT  DISULFID  3903  3936  BY SIMILARITY.
FT  DISULFID  3936  3969  BY SIMILARITY.
FT  DISULFID  3969  4002  BY SIMILARITY.
FT  DISULFID  4002  4035  BY SIMILARITY.
FT  DISULFID  4035  4068  BY SIMILARITY.
FT  DISULFID  4068  4101  BY SIMILARITY.
FT  DISULFID  4101  4134  BY SIMILARITY.
FT  DISULFID  4134  4167  BY SIMILARITY.
FT  DISULFID  4167  4200  BY SIMILARITY.
FT  DISULFID  4200  4233  BY SIMILARITY.
FT  DISULFID  4233  4266  BY SIMILARITY.
FT  DISULFID  4266  4299  BY SIMILARITY.
FT  DISULFID  4299  4332  BY SIMILARITY.
FT  DISULFID  4332  4365  BY SIMILARITY.
FT  DISULFID  4365  4398  BY SIMILARITY.
FT  DISULFID  4398  4431  BY SIMILARITY.
FT  DISULFID  4431  4464  BY SIMILARITY.
FT  DISULFID  4464  4497  BY SIMILARITY.
FT  DISULFID  4497  4530  BY SIMILARITY.
FT  DISULFID  4530  4563  BY SIMILARITY.
FT  DISULFID  4563  4596  BY SIMILARITY.
FT  DISULFID  4596  4629  BY SIMILARITY.
FT  DISULFID  4629  4662  BY SIMILARITY.
FT  DISULFID  4662  4695  BY SIMILARITY.
FT  DISULFID  4695  4728  BY SIMILARITY.
FT  DISULFID  4728  4761  BY SIMILARITY.
FT  DISULFID  4761  4794  BY SIMILARITY.
FT  DISULFID  4794  4827  BY SIMILARITY.
FT  DISULFID  4827  4860  BY SIMILARITY.
FT  DISULFID  4860  4893  BY SIMILARITY.
FT  DISULFID  4893  4926  BY SIMILARITY.
FT  DISULFID  4926  4959  BY SIMILARITY.
FT  DISULFID  4959  4992  BY SIMILARITY.
FT  DISULFID  4992  5025  BY SIMILARITY.
FT  DISULFID  5025  5058  BY SIMILARITY.
FT  DISULFID  5058  5091  BY SIMILARITY.
FT  DISULFID  5091  5124  BY SIMILARITY.
FT  DISULFID  5124  5157  BY SIMILARITY.
FT  DISULFID  5157  5190  BY SIMILARITY.
FT  DISULFID  5190  5223  BY SIMILARITY.
FT  DISULFID  5223  5256  BY SIMILARITY.
FT  DISULFID  5256  5289  BY SIMILARITY.
FT  DISULFID  5289  5322  BY SIMILARITY.
FT  DISULFID  5322  5355  BY SIMILARITY.
FT  DISULFID  5355  5388  BY SIMILARITY.
FT  DISULFID  5388  5421  BY SIMILARITY.
FT  DISULFID  5421  5454  BY SIMILARITY.
FT  DISULFID  5454  5487  BY SIMILARITY.
FT  DISULFID  5487  5520  BY SIMILARITY.
FT  DISULFID  5520  5553  BY SIMILARITY.
FT  DISULFID  5553  5586  BY SIMILARITY.
FT  DISULFID  5586  5619  BY SIMILARITY.
FT  DISULFID  5619  5652  BY SIMILARITY.
FT  DISULFID  5652  5685  BY SIMILARITY.
FT  DISULFID  5685  5718  BY SIMILARITY.
FT  DISULFID  5718  5751  BY SIMILARITY.
FT  DISULFID  5751  5784  BY SIMILARITY.
FT  DISULFID  5784  5817  BY SIMILARITY.
FT  DISULFID  5817  5850  BY SIMILARITY.
FT  DISULFID  5850  5883  BY SIMILARITY.
FT  DISULFID  5883  5916  BY SIMILARITY.
FT  DISULFID  5916  5949  BY SIMILARITY.
FT  DISULFID  5949  5982  BY SIMILARITY.
FT  DISULFID  5982  6015  BY SIMILARITY.
FT  DISULFID  6015  6048  BY SIMILARITY.
FT  DISULFID  6048  6081  BY SIMILARITY.
FT  DISULFID  6081  6114  BY SIMILARITY.
FT  DISULFID  6114  6147  BY SIMILARITY.
FT  DISULFID  6147  6180  BY SIMILARITY.
FT  DISULFID  6180  6213  BY SIMILARITY.
FT  DISULFID  6213  6246  BY SIMILARITY.
FT  DISULFID  6246  6279  BY SIMILARITY.
FT  DISULFID  6279  6312  BY SIMILARITY.
FT  DISULFID  6312  6345  BY SIMILARITY.
FT  DISULFID  6345  6378  BY SIMILARITY.
FT  DISULFID  6378  6411  BY SIMILARITY.
FT  DISULFID  6411  6444  BY SIMILARITY.
FT  DISULFID  6444  6477  BY SIMILARITY.
FT  DISULFID  6477  6510  BY SIMILARITY.
FT  DISULFID  6510  6543  BY SIMILARITY.
FT  DISULFID  6543  6576  BY SIMILARITY.
FT  DISULFID  6576  6609  BY SIMILARITY.
FT  DISULFID  6609  6642  BY SIMILARITY.
FT  DISULFID  6642  6675  BY SIMILARITY.
FT  DISULFID  6675  6708  BY SIMILARITY.
FT  DISULFID  6708  6741  BY SIMILARITY.
FT  DISULFID  6741  6774  BY SIMILARITY.
FT  DISULFID  6774  6807  BY SIMILARITY.
FT  DISULFID  6807  6840  BY SIMILARITY.
FT  DISULFID  6840  6873  BY SIMILARITY.
FT  DISULFID  6873  6906  BY SIMILARITY.
FT  DISULFID  6906  6939  BY SIMILARITY.
FT  DISULFID  6939  6972  BY SIMILARITY.
FT  DISULFID  6972  7005  BY SIMILARITY.
FT  DISULFID  7005  7038  BY SIMILARITY.
FT  DISULFID  7038  7071  BY SIMILARITY.
FT  DISULFID  7071  7104  BY SIMILARITY.
FT  DISULFID  7104  7137  BY SIMILARITY.
FT  DISULFID  7137  7170  BY SIMILARITY.
FT  DISULFID  7170  7203  BY SIMILARITY.
FT  DISULFID  7203  7236  BY SIMILARITY.
FT  DISULFID  7236  7269  BY SIMILARITY.
FT  DISULFID  7269  7302  BY SIMILARITY.
FT  DISULFID  7302  7335  BY SIMILARITY.
FT  DISULFID  7335  7368  BY SIMILARITY.
FT  DISULFID  7368  7401  BY SIMILARITY.
FT  DISULFID  7401  7434  BY SIMILARITY.
FT  DISULFID  7434  7467  BY SIMILARITY.
FT  DISULFID  7467  7500  BY SIMILARITY.
FT  DISULFID  7500  7533  BY SIMILARITY.
FT  DISULFID  7533  7566  BY SIMILARITY.
FT  DISULFID  7566  7599  BY SIMILARITY.
FT  DISULFID  7599  7632  BY SIMILARITY.
FT  DISULFID  7632  7665  BY SIMILARITY.
FT  DISULFID  7665  7698  BY SIMILARITY.
FT  DISULFID  7698  7731  BY SIMILARITY.
FT  DISULFID  7731  7764  BY SIMILARITY.
FT  DISULFID  7764  7797  BY SIMILARITY.
FT  DISULFID  7797  7830  BY SIMILARITY.
FT  DISULFID  7830  7863  BY SIMILARITY.
FT  DISULFID  7863  7896  BY SIMILARITY.
FT  DISULFID  7896  7929  BY SIMILARITY.
FT  DISULFID  7929  7962  BY SIMILARITY.
FT  DISULFID  7962  7995  BY SIMILARITY.
FT  DISULFID  7995  8028  BY SIMILARITY.
FT  DISULFID  8028  8061  BY SIMILARITY.
FT  DISULFID  8061  8094  BY SIMILARITY.
FT  DISULFID  8094  8127  BY SIMILARITY.
FT  DISULFID  8127  8160  BY SIMILARITY.
FT  DISULFID  8160  8193  BY SIMILARITY.
FT  DISULFID  8193  8226  BY SIMILARITY.
FT  DISULFID  8226  8259  BY SIMILARITY.
FT  DISULFID  8259  8292  BY SIMILARITY.
FT  DISULFID  8292  8325  BY SIMILARITY.
FT  DISULFID  8325  8358  BY SIMILARITY.
FT  DISULFID  8358  8391  BY SIMILARITY.
FT  DISULFID  8391  8424  BY SIMILARITY.
FT  DISULFID  8424  8457  BY SIMILARITY.
FT  DISULFID  8457  8490  BY SIMILARITY.
FT  DISULFID  8490  8523  BY SIMILARITY.
FT  DISULFID  8523  8556  BY SIMILARITY.
FT  DISULFID  8556  8589  BY SIMILARITY.
FT  DISULFID  8589  8622  BY SIMILARITY.
FT  DISULFID  8622  8655  BY SIMILARITY.
FT  DISULFID  8655  8688  BY SIMILARITY.
FT  DISULFID  8688  8721  BY SIMILARITY.
FT  DISULFID  8721  8754  BY SIMILARITY.
FT  DISULFID  8754  8787  BY SIMILARITY.
FT  DISULFID  8787  8820  BY SIMILARITY.
FT  DISULFID  8820  8853  BY SIMILARITY.
FT  DISULFID  8853  8886  BY SIMILARITY.
FT  DISULFID  8886  8919  BY SIMILARITY.
FT  DISULFID  8919  8952  BY SIMILARITY.
FT  DISULFID  8952  8985  BY SIMILARITY.
FT  DISULFID  8985  9018  BY SIMILARITY.
FT  DISULFID  9018  9051  BY SIMILARITY.
FT  DISULFID  9051  9084  BY SIMILARITY.
FT  DISULFID  9084  9117  BY SIMILARITY.
FT  DISULFID  9117  9150  BY SIMILARITY.
FT  DISULFID  9150  9183  BY SIMILARITY.
FT  DISULFID  9183  9216  BY SIMILARITY.
FT  DISULFID  9216  9249  BY SIMILARITY.
FT  DISULFID  9249  9282  BY SIMILARITY.
FT  DISULFID  9282  9315  BY SIMILARITY.
FT  DISULFID  9315  9348  BY SIMILARITY.
FT  DISULFID  9348  9381  BY SIMILARITY.
FT  DISULFID  9381  9414  BY SIMILARITY.
FT  DISULFID  9414  9447  BY SIMILARITY.
FT  DISULFID  9447  9480  BY SIMILARITY.
FT  DISULFID  9480  9513  BY SIMILARITY.
FT  DISULFID  9513  9546  BY SIMILARITY.
FT  DISULFID  9546  9579  BY SIMILARITY.
FT  DISULFID  9579  9612  BY SIMILARITY.
FT  DISULFID  9612  9645  BY SIMILARITY.
FT  DISULFID  9645  9678  BY SIMILARITY.
FT  DISULFID  9678  9711  BY SIMILARITY.
FT  DISULFID  9711  9744  BY SIMILARITY.
FT  DISULFID  9744  9777  BY SIMILARITY.
FT  DISULFID  9777  9810  BY SIMILARITY.
FT  DISULFID  9810  9843  BY SIMILARITY.
FT  DISULFID  9843  9876  BY SIMILARITY.
FT  DISULFID  9876  9909  BY SIMILARITY.
FT  DISULFID  9909  9942  BY SIMILARITY.
FT  DISULFID  9942  9975  BY SIMILARITY.
FT  DISULFID  9975  10008  BY SIMILARITY.
FT  DISULFID  10008  10041  BY SIMILARITY.
FT  DISULFID  10041  10074  BY SIMILARITY.
FT  DISULFID  10074  10107  BY SIMILARITY.
FT  DISULFID  10107  10140  BY SIMILARITY.
FT  DISULFID  10140  10173  BY SIMILARITY.
FT  DISULFID  10173  10206  BY SIMILARITY.
FT  DISULFID  10206  10239  BY SIMILARITY.
FT  DISULFID  10239  10272  BY SIMILARITY.
FT  DISULFID  10272  10305  BY SIMILARITY.
FT  DISULFID  10305  10338  BY SIMILARITY.
FT  DISULFID  10338  10371  BY SIMILARITY.
FT  DISULFID  10371  10404  BY SIMILARITY.
FT  DISULFID  10404  10437  BY SIMILARITY.
FT  DISULFID  10437  10470  BY SIMILARITY.
FT  DISULFID  10470  10503  BY SIMILARITY.
FT  DISULFID  10503  10536  BY SIMILARITY.
FT  DISULFID  10536  10569  BY SIMILARITY.
FT  DISULFID  10569  10602  BY SIMILARITY.
FT  DISULFID  10602  10635  BY SIMILARITY.
FT  DISULFID  10635  10668  BY SIMILARITY.
FT  DISULFID  10668  10701  BY SIMILARITY.
FT  DISULFID  10701  10734  BY SIMILARITY.
FT  DISULFID  10734  10767  BY SIMILARITY.
FT  DISULFID  10767  10800  BY SIMILARITY.
FT  DISULFID  10800  10833  BY SIMILARITY.
FT  DISULFID  10833  10866  BY SIMILARITY.
FT  DISULFID  10866  10899  BY SIMILARITY.
FT  DISULFID  10899  10932  BY SIMILARITY.
FT  DISULFID  10932  10965  BY SIMILARITY.
FT  DISULFID  10965  10998  BY SIMILARITY.
FT  DISULFID  10998  11031  BY SIMILARITY.
FT  DISULFID  11031  11064  BY SIMILARITY.
FT  DISULFID  11064  11097  BY SIMILARITY.
FT  DISULFID  11097  11130  BY SIMILARITY.
FT  DISULFID  11130  11163  BY SIMILARITY.
FT  DISULFID  11163  11196  BY SIMILARITY.
FT  DISULFID  11196  11229  BY SIMILARITY.
FT  DISULFID  11229  11262  BY SIMILARITY.
FT  DISULFID  11262  11295  BY SIMILARITY.
FT  DISULFID  11295  11328  BY SIMILARITY.
FT  DISULFID  11328  11361  BY SIMILARITY.
FT  DISULFID  11361  11394  BY SIMILARITY.
FT  DISULFID  11394  11427  BY SIMILARITY.
FT  DISULFID  11427  11460  BY SIMILARITY.
FT  DISULFID  11460  11493  BY SIMILARITY.
FT  DISULFID  11493  11526  BY SIMILARITY.
FT  DISULFID  11526  11559  BY SIMILARITY.
FT  DISULFID  11559  11592  BY SIMILARITY.
FT  DISULFID  11592  11625  BY SIMILARITY.
FT  DISULFID  11625  11658  BY SIMILARITY.
FT  DISULFID  11658  11691  BY SIMILARITY.
FT  DISULFID  11691  11724  BY SIMILARITY.
FT  DISULFID  11724  11757  BY SIMILARITY.
FT  DISULFID  11757  11790  BY SIMILARITY.
FT  DISULFID  11790  11823  BY SIMILARITY.
FT  DISULFID  11823  11856  BY SIMILARITY.
FT  DISULFID  11856  11889  BY SIMILARITY.
FT  DISULFID  11889  11922  BY SIMILARITY.
FT  DISULFID  11922  11955  BY SIMILARITY.
FT  DISULFID  11955  11988  BY SIMILARITY.
FT  DISULFID  11988  12021  BY SIMILARITY.
FT  DISULFID  12021  12054  BY SIMILARITY.
FT  DISULFID  12054  12087  BY SIMILARITY.
FT  DISULFID  12087  12120  BY SIMILARITY.
FT  DISULFID  12120  12153  BY SIMILARITY.
FT  DISULFID  12153  12186  BY SIMILARITY.
FT  DISULFID  12186  12219  BY SIMILARITY.
FT  DISULFID  12219  12252  BY SIMILARITY.
FT  DISULFID  12252  12285  BY SIMILARITY.
FT  DISULFID  12285  12318  BY SIMILARITY.
FT  DISULFID  12318  12351  BY SIMILARITY.
FT  DISULFID  12351  12384  BY SIMILARITY.
FT  DISULFID  12384  12417  BY SIMILARITY.
FT  DISULFID  12417  12450  BY SIMILARITY.
FT  DISULFID  12450  12483  BY SIMILARITY.
FT  DISULFID  12483  12516  BY SIMILARITY.
FT  DISULFID  12516  12549  BY SIMILARITY.
FT  DISULFID  12549  12582  BY SIMILARITY.
FT  DISULFID  12582  12615  BY SIMILARITY.
FT  DISULFID  12615  12648  BY SIMILARITY.
FT  DISULFID  12648  12681  BY SIMILARITY.
FT  DISULFID  12681  12714  BY SIMILARITY.
FT  DISULFID  12714  12747  BY SIMILARITY.
FT  DISULFID  12747  12780  BY SIMILARITY.
FT  DISULFID  12780  12813  BY SIMILARITY.
FT  DISULFID  12813  12846  BY SIMILARITY.
FT  DISULFID  12846  12879  BY SIMILARITY.
FT  DISULFID  12879  12912  BY SIMILARITY.
FT  DISULFID  12912  12945  BY SIMILARITY.
FT  DISULFID  12945  12978  BY SIMILARITY.
FT  DISULFID  12978  13011  BY SIMILARITY.
FT  DISULFID  13011  13044  BY SIMILARITY.
FT  DISULFID  13044  13077  BY SIMILARITY.
FT  DISULFID  13077  13110  BY SIMILARITY.
FT  DISULFID  13110  13143  BY SIMILARITY.
FT  DISULFID  13143  13176  BY SIMILARITY.
FT  DISULFID  13176  13209  BY SIMILARITY.
FT  DISULFID  13209  13242  BY SIMILARITY.
FT  DISULFID  13242  13275  BY SIMILARITY.
FT  DISULFID  13275  13308  BY SIMILARITY.
FT  DISULFID  13308  13341  BY SIMILARITY.
FT  DISULFID  13341  13374  BY SIMILARITY.
FT  DISULFID  13374  13407  BY SIMILARITY.
FT  DISULFID  13407  13440  BY SIMILARITY.
FT  DISULFID  13440  13473  BY SIMILARITY.
FT  DISULFID  13473  13506  BY SIMILARITY.
FT  DISULFID  13506  13539  BY SIMILARITY.
FT  DISULFID  13539  13572  BY SIMILARITY.
FT  DISULFID  13572  13605  BY SIMILARITY.
FT  DISULFID  13605  13638  BY SIMILARITY.
FT  DISULFID  13638  13671  BY SIMILARITY.
FT  DISULFID  13671  13704  BY SIMILARITY.
FT  DISULFID  13704  13737  BY SIMILARITY.
FT  DISULFID  13737  13770  BY SIMILARITY.
FT  DISULFID  13770  13803  BY SIMILARITY.
FT  DISULFID  13803  13836  BY SIMILARITY.
FT  DISULFID  13836  13869  BY SIMILARITY.
FT  DISULFID  13869  13902  BY SIMILARITY.
FT  DISULFID  13902  13935  BY SIMILARITY.
FT  DISULFID  13935  13968  BY SIMILARITY.
FT  DISULFID  13968  14001  BY SIMILARITY.
FT  DISULFID  14001  14034  BY SIMILARITY.
FT  DISULFID  14034  14067  BY SIMILARITY.
FT  DISULFID  14067  14100  BY SIMILARITY.
FT  DISULFID  14100  14133  BY SIMILARITY.
FT  DISULFID  14133  14166  BY SIMILARITY.
FT  DISULFID  14166  14199  BY SIMILARITY.
FT  DISULFID  14199  14232  BY SIMILARITY.
FT  DISULFID  14232  14265  BY SIMILARITY.
FT  DISULFID  14265  14298  BY SIMILARITY.
FT  DISULFID  14298  14331  BY SIMILARITY.
FT  DISULFID  14331  14364  BY SIMILARITY.
FT  DISULFID  14364  14397  BY SIMILARITY.
FT  DISULFID  14397  14430  BY SIMILARITY.
FT  DISULFID  14430  14463  BY SIMILARITY.
FT  DISULFID  14463  14496  BY SIMILARITY.
FT  DISULFID  14496  14529  BY SIMILARITY.
FT  DISULFID  14529  14562  BY SIMILARITY.
FT  DISULFID  14562  14595  BY SIMILARITY.
FT  DISULFID  14595  14628  BY SIMILARITY.
FT  DISULFID  14628  14661  BY SIMILARITY.
FT  DISULFID  14661  14694  BY SIMILARITY.
FT  DISULFID  14694  14727  BY SIMILARITY.
FT  DISULFID  14727  14760  BY SIMILARITY.
FT  DISULFID  14760  14793  BY SIMILARITY.
FT  DISULFID  14793  14826  BY SIMILARITY.
FT  DISULFID  14826  14859  BY SIMILARITY.
FT  DISULFID  14859  14892  BY SIMILARITY.
FT  DISULFID  14892  14925  BY SIMILARITY.
FT  DISULFID  14925  14958  BY SIMILARITY.
FT  DISULFID  14958  14991  BY SIMILARITY.
FT  DISULFID  14991  15024  BY SIMILARITY.
FT  DISULFID  15024  15057  BY SIMILARITY.
FT  DISULFID  15057  15090  BY SIMILARITY.
FT  DISULFID  15090  15123  BY SIMILARITY.
FT  DISULFID  15123  15156  BY SIMILARITY.
FT  DISULFID  15156  15189  BY SIMILARITY.
FT  DISULFID  15189  15222  BY SIMILARITY.
FT  DISULFID  15222  15255  BY SIMILARITY.
FT  DISULFID  15255  15288  BY SIMILARITY.
FT  DISULFID  15288  15321  BY SIMILARITY.
FT  DISULFID  15321  15354  BY SIMILARITY.
FT  DISULFID  15354  15387  BY SIMILARITY.
FT  DISULFID  15387  15420  BY SIMILARITY.
FT  DISULFID  15420  15453  BY SIMILARITY.
FT  DISULFID  15453  15486  BY SIMILARITY.
FT  DISULFID  15486  15519  BY SIMILARITY.
FT  DISULFID  15519  15552  BY SIMILARITY.
FT  DISULFID  15552  15585  BY SIMILARITY.
FT  DISULFID  15585  15618  BY SIMILARITY.
FT  DISULFID  15618  15651  BY SIMILARITY.
FT  DISULFID  15651  15684  BY SIMILARITY.
FT  DISULFID  15684  15717  BY SIMILARITY.
FT  DISULFID  15717  15750  BY SIMILARITY.
FT  DISULFID  15750  15783  BY SIMILARITY.
FT  DISULFID  15783  15816  BY SIMILARITY.
FT  DISULFID  15816  15849  BY SIMILARITY.
FT  DISULFID  15849  15882  BY SIMILARITY.
FT  DISULFID  15882  15915  BY SIMILARITY.
FT  DISULFID  15915  15948  BY SIMILARITY.
FT  DISULFID  15948  15981  BY SIMILARITY.
FT  DISULFID  15981  16014  BY SIMILARITY.
FT  DISULFID  16014  16047  BY SIMILARITY.
FT  DISULFID  16047  1608
```

DR Pfam; PF00008; EGF; 33.
 DR PIRSE; PIRSF00066; notch; 3.
 DR PRINTS; PR00010; EGFLOOD.
 DR PRINTS; PR00011; EGFAMININ.
 DR SMART; SM00248; ANK; 6.
 DR SMART; SM00179; EGF_CA; 20.
 DR SMART; SM00004; NL; 3.
 DR PROSITE; PS00297; ANK_REPEAT; 1.
 DR PROSITE; PS00088; ANK_REPEAT; 4.
 DR PROSITE; PS00010; ASX_HYDROXYL; 18.
 DR PROSITE; PS00022; EGF_1; 33.
 DR PROSITE; PS01186; EGF_2; 26.
 DR PROSITE; PS00026; EGF_3; 34.
 DR PROSITE; PS01187; EGF_CA; 16.
 DR Receptor; Transcription regulation; Activator; Differentiation;
 DR Developmental protein; Repeat; EGF-like domain;
 DR Transmembrane; Glycoprotein; Signal; Phosphorylation.
 KW SIGNAL.
 FT 1 40
 FT CHAIN 41 2319
 FT CHAIN 1631 2319
 FT CHAIN 1664 2319
 FT DOMAIN 41 1645
 FT TRANSMEM 1646 1666
 FT DOMAIN 1667 2319
 FT DOMAIN 41 79
 FT DOMAIN 121 158
 FT DOMAIN 160 197
 FT DOMAIN 199 236
 FT DOMAIN 238 274
 FT DOMAIN 276 314
 FT DOMAIN 316 352
 FT DOMAIN 353 391
 FT DOMAIN 393 431
 FT DOMAIN 433 469
 FT DOMAIN 471 507
 FT DOMAIN 509 545
 FT DOMAIN 547 582
 FT DOMAIN 620 657
 FT DOMAIN 622 657
 FT DOMAIN 659 695
 FT DOMAIN 697 732
 FT DOMAIN 736 772
 FT DOMAIN 773 810
 FT DOMAIN 812 849
 FT DOMAIN 851 887
 FT DOMAIN 889 924
 FT DOMAIN 926 962
 FT DOMAIN 964 1000
 FT DOMAIN 1002 1036
 FT DOMAIN 1038 1084
 FT DOMAIN 1086 1122
 FT DOMAIN 1124 1160
 FT DOMAIN 1160 1205
 FT DOMAIN 1207 1246
 FT DOMAIN 1248 1289
 FT DOMAIN 1291 1327
 FT DOMAIN 1337 1375
 FT REPEAT 1384 1420
 FT REPEAT 1425 1461
 FT REPEAT 1469 1503
 FT REPEAT 1840 1869
 FT REPEAT 1873 1903
 FT REPEAT 1907 1936
 FT REPEAT 1940 1969
 FT REPEAT 2002 2024
 FT SITE 1573 1574
 FT DISULFID 44 56

POTENTIAL.
 NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 3.
 NOTCH EXTRACELLULAR TRUNCATION (BY
 SIMILARITY).
 NOTCH INTRACELLULAR DOMAIN (BY
 SIMILARITY).
 EXTRACELLULAR (POTENTIAL).
 POTENTIAL.
 CYTOPLASMIC (POTENTIAL).
 EGF-LIKE 1.
 EGF-LIKE 2.
 EGF-LIKE 3.
 EGF-LIKE 4.
 EGF-LIKE 5.
 EGF-LIKE 6.
 EGF-LIKE 7.
 EGF-LIKE 8.
 EGF-LIKE 9.
 EGF-LIKE 10.
 EGF-LIKE 11.
 EGF-LIKE 12.
 EGF-LIKE 13.
 EGF-LIKE 14.
 EGF-LIKE 15.
 EGF-LIKE 16.
 EGF-LIKE 17.
 EGF-LIKE 18.
 EGF-LIKE 19.
 EGF-LIKE 20.
 EGF-LIKE 21.
 EGF-LIKE 22.
 EGF-LIKE 23.
 EGF-LIKE 24.
 EGF-LIKE 25.
 EGF-LIKE 26.
 EGF-LIKE 27.
 EGF-LIKE 28.
 EGF-LIKE 29.
 EGF-LIKE 30.
 EGF-LIKE 31.
 EGF-LIKE 32.
 EGF-LIKE 33.
 EGF-LIKE 34.
 LIN/NOTCH 1.
 LIN/NOTCH 2.
 LIN/NOTCH 3.
 ANK 1.
 ANK 2.
 ANK 3.
 ANK 4.
 ANK 5.
 CLEAVAGE (BY FUJIN-LIKE PROTEASE) (BY
 SIMILARITY).
 BY SIMILARITY.

FT DISULFID 50 67 BY SIMILARITY.
 FT DISULFID 69 78 BY SIMILARITY.
 FT DISULFID 84 95 BY SIMILARITY.
 FT DISULFID 89 108 BY SIMILARITY.
 FT DISULFID 110 119 BY SIMILARITY.
 FT DISULFID 115 125 BY SIMILARITY.
 FT DISULFID 130 136 BY SIMILARITY.
 FT DISULFID 130 146 BY SIMILARITY.
 FT DISULFID 148 157 BY SIMILARITY.
 FT DISULFID 164 176 BY SIMILARITY.
 FT DISULFID 170 185 BY SIMILARITY.
 FT DISULFID 187 196 BY SIMILARITY.
 FT DISULFID 203 214 BY SIMILARITY.
 FT DISULFID 208 224 BY SIMILARITY.
 FT DISULFID 226 235 BY SIMILARITY.
 FT DISULFID 242 253 BY SIMILARITY.
 FT DISULFID 247 262 BY SIMILARITY.
 FT DISULFID 264 273 BY SIMILARITY.
 FT DISULFID 280 293 BY SIMILARITY.
 FT DISULFID 287 302 BY SIMILARITY.
 FT DISULFID 304 313 BY SIMILARITY.
 FT DISULFID 320 331 BY SIMILARITY.
 FT DISULFID 325 340 BY SIMILARITY.
 FT DISULFID 342 351 BY SIMILARITY.
 FT DISULFID 357 368 BY SIMILARITY.
 FT DISULFID 362 379 BY SIMILARITY.
 FT DISULFID 381 390 BY SIMILARITY.
 FT DISULFID 397 410 BY SIMILARITY.
 FT DISULFID 404 419 BY SIMILARITY.
 FT DISULFID 421 430 BY SIMILARITY.
 FT DISULFID 437 448 BY SIMILARITY.
 FT DISULFID 442 457 BY SIMILARITY.
 FT DISULFID 459 468 BY SIMILARITY.
 FT DISULFID 475 486 BY SIMILARITY.
 FT DISULFID 480 495 BY SIMILARITY.
 FT DISULFID 497 506 BY SIMILARITY.
 FT DISULFID 513 524 BY SIMILARITY.
 FT DISULFID 518 533 BY SIMILARITY.
 FT DISULFID 535 544 BY SIMILARITY.
 FT DISULFID 551 561 BY SIMILARITY.

Query Match 59.7%; Score 83; DB 1; Length 2319;
 Best local Similarity 41.9%; Pred. No. 0.00011;
 Matches 13; Conservative 2; Mismatches 16; Indels 0; Gaps 0;
 QY 1 CDYXXCXXXXXGXCXCCNNAACXXDXDC 31
 Db 1389 CPRAACQAKRGDQNCDBRECNPGCGWDGDC 1419

RESULT 11
 ID NTCS HUMAN STANDARD; PRT; 2321 AA.
 AC G9M47; Q9UEB3; Q9UEB3; Q9Y6L8;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DE Neurogenic locus notch homolog protein 3 precursor (Notch 3).
 GN NOTCH3.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97032728; PubMed=8878479;
 RA Joutel A., Cornejo C., Ducros A., Vahedi K., Chabriat H., Mouton P.,
 RA Alamiouich S., Domenga V., Cecillon M., Marchal E., Maciazek J.,
 RA Vaysiere C., Cruaud C., Cabanis E.-A., Ruchoux M.M., Weissenbach J.,
 RA Bach J.-F., Bousset M.-G., Tourner-Lasserre E.,
 RT "Notch3 mutations in CADASIL, a hereditary adult-onset condition
 causing stroke and dementia.",
 RL Nature 383:707-710 (1996).
 RN [2]

RP SEQUENCE FROM N.A.
 RA Gmel M., Artavanis-Tsakonas S.;
 RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Lamedin J.E., McCready P.M., Skowronski E., Adamson A.W.,
 RA Burkhart-Schultz K., Gordon L., Kyle A., Ramirez M., Stillingen S.,
 RA Phan H., Velasco N., Gaines J., Dangnan L., Poundstone P.,
 RA Christensen M., Georgescu A., Avila J., Liu S., Attix C., Andreise T.,
 RA Tranheim M., Amico-Keller G., Coefield J., Duarte S., Lucas S.,
 RA Bruce R., Thomas P., Quan G., Krommiller B., Arellano A.,
 RA Montomery M., Ow D., Nolan M., Trong S., Kobayashi A., Olsen A.S.,
 RA Carreno A.V.;
 RT "Sequence analysis of an 1.5 Mb olfactory receptor (OLFR) cluster in
 RT 19p13.1";
 RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP VARIANTS CADASIL TYR-49; CYS-71; CYS-90; CYS-110; CYS-133; CYS-141;
 RP ARG-146; CYS-153; CYS-169; CYS-171; CYS-182; ARG-185; SER-212;
 RP GLY-222; TYR-224; CYS-258; TYR-542; CYS-558; CYS-578; CYS-728;
 RP CYS-985; CYS-1006; CYS-1031; CYS-1231 AND ARG-1261, AND VARIANTS
 RP ARG-170; LEU-496; GLN-1133; MET-1183 AND ALA-2223.
 RX MEDLINE=98049753; PubMed=9388398;
 RA Younel A., Vahedi K., Corpechot C., Troesch A., Chabrier H.,
 RA Vaysiere C., Chuad C., Maciazek J., Weissbach J., Bousser M.-G.,
 RA Bach J.-F., Tournier-Lasserre E.;
 RT "Strong clustering and stereotyped nature of Notch3 mutations in
 RT CADASIL patients";
 RL Lancet 350:1511-1515(1997).
 RN [5]
 RP VARIANT CADASIL 114-GLY-PRO-120 DEL.
 RX MEDLINE=20264473; PubMed=10802807;
 RA Couteil A., Chabrier H., Vahedi K., Domenga V., Vaysiere C.,
 RA Ruchoux M.M., Lucas C., Lays D., Bousser M.-G., Tournier-Lasserre E.;
 RT "Splice site mutation causing a seven amino acid Notch3 in-frame
 RT deletion in CADASIL";
 RL Neurology 54:1874-1875(2000).
 RN [6]
 RP IDENTIFICATION OF LIGANDS.
 RX MEDLINE=99180765; PubMed=10079256;
 RA Gray G.E., Mann R.S., Mitsiadis E., Henrique D., Carcangu M.-L.,
 RA Banks A., Leiman J., Ward D., Ish-Horowitz D., Artavanis-Tsakonas S.;
 RL Am. J. Pathol. 154:785-794(1999).
 CC -1- FUNCTION: Functions as a receptor for membrane-bound ligands
 CC Jagged1, Jagged2 and Delta to regulate cell-fate determination.
 CC Upon ligand activation through the released notch intracellular
 CC domain (NICD) it forms a transcriptional activator complex with
 CC RBP-J kappa and activates genes of the enhancer of split locus.
 CC Affects the implementation of differentiation, proliferation and
 CC apoptotic programs (By similarity).
 CC -1- SUBUNIT: Heterodimer of a C-terminal fragment N(TW) and a N-
 CC terminal fragment N(BC) which are probably linked by disulfide
 CC bonds (By similarity).
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Following
 CC proteolytical processing NICD is translocated to the nucleus.
 CC -1- TISSUE SPECIFICITY: Ubiquitously expressed in fetal and adult
 CC tissues.
 CC -1- PTM: Synthesized in the endoplasmic reticulum as an inactive form
 CC which is proteolytically cleaved by a furin-like convertase to yield
 CC trans-golgi network before it reaches the plasma membrane to yield
 CC an active, ligand-accessible form. Cleavage results in a C-
 CC terminal fragment N(TW) and a N-terminal fragment N(BC). Following
 CC ligand binding, it is cleaved by TNF-alpha converting enzyme
 CC (TACE) to yield a membrane-associated intermediate fragment called
 CC notch extracellular truncation (NEXT). This fragment is then
 CC cleaved by presenilin dependent gamma-secretase to release a
 CC notch-derived peptide containing the intracellular domain (NICD)
 CC from the membrane (By similarity).
 CC -1- PTM: Phosphorylated (By similarity).
 CC -1- DISEASE: Defects in NOTCH3 are associated with cerebral autosomal
 CC dominant arteriopathy with subcortical infarcts and
 CC leukoencephalopathy (CADASIL) [MIM:125310]. CADASIL causes a type

CC of stroke and dementia of which key features include recurrent
 CC subcortical ischemic events and vascular dementia.
 CC -1- SIMILARITY: Belongs to the NOTCH family.
 CC -1- SIMILARITY: Contains 34 EGF-like domains.
 CC -1- SIMILARITY: Contains 3 Lin/Notch repeats.
 CC -1- SIMILARITY: Contains 5 ANK repeats.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U97669; AAB91371.1; -;
 CC EMBL; AF058900; AAC14346.1; -;
 CC EMBL; AF058881; AAC14346.1; JOINED.
 CC EMBL; AF058882; AAC14346.1; JOINED.
 CC EMBL; AF058883; AAC14346.1; JOINED.
 CC EMBL; AF058884; AAC14346.1; JOINED.
 CC EMBL; AF058885; AAC14346.1; JOINED.
 CC EMBL; AF058886; AAC14346.1; JOINED.
 CC EMBL; AF058887; AAC14346.1; JOINED.
 CC EMBL; AF058888; AAC14346.1; JOINED.
 CC EMBL; AF058889; AAC14346.1; JOINED.
 CC EMBL; AF058890; AAC14346.1; JOINED.
 CC EMBL; AF058891; AAC14346.1; JOINED.
 CC EMBL; AF058892; AAC14346.1; JOINED.
 CC EMBL; AF058893; AAC14346.1; JOINED.
 CC EMBL; AF058894; AAC14346.1; JOINED.
 CC EMBL; AF058895; AAC14346.1; JOINED.
 CC EMBL; AF058896; AAC14346.1; JOINED.
 CC EMBL; AF058897; AAC14346.1; JOINED.
 CC EMBL; AF058898; AAC14346.1; JOINED.
 CC EMBL; AF058899; AAC14346.1; JOINED.
 CC EMBL; AF058900; AAC14346.1; JOINED.
 CC EMBL; AF058901; AAC14346.1; JOINED.
 CC EMBL; AF058902; AAC14346.1; JOINED.
 CC EMBL; AF058903; AAC14346.1; JOINED.
 CC EMBL; AF058904; AAC14346.1; JOINED.
 CC EMBL; AF058905; AAC14346.1; JOINED.
 CC EMBL; AF058906; AAC14346.1; JOINED.
 CC EMBL; AF058907; AAC14346.1; JOINED.
 CC EMBL; AF058908; AAC14346.1; JOINED.
 CC EMBL; AF058909; AAC14346.1; JOINED.
 CC EMBL; AF058910; AAC14346.1; JOINED.
 CC EMBL; AF058911; AAC14346.1; JOINED.
 CC EMBL; AF058912; AAC14346.1; JOINED.
 CC EMBL; AF058913; AAC14346.1; JOINED.
 CC EMBL; AF058914; AAC14346.1; JOINED.
 CC EMBL; AF058915; AAC14346.1; JOINED.
 CC EMBL; AF058916; AAC14346.1; JOINED.
 CC EMBL; AF058917; AAC14346.1; JOINED.
 CC EMBL; AF058918; AAC14346.1; JOINED.
 CC EMBL; AF058919; AAC14346.1; JOINED.
 CC EMBL; AF058920; AAC14346.1; JOINED.
 CC EMBL; AF058921; AAC14346.1; JOINED.
 CC EMBL; AF058922; AAC14346.1; JOINED.
 CC EMBL; AF058923; AAC14346.1; JOINED.
 CC EMBL; AF058924; AAC14346.1; JOINED.
 CC EMBL; AF058925; AAC14346.1; JOINED.
 CC EMBL; AF058926; AAC14346.1; JOINED.
 CC EMBL; AF058927; AAC14346.1; JOINED.
 CC EMBL; AF058928; AAC14346.1; JOINED.
 CC EMBL; AF058929; AAC14346.1; JOINED.
 CC EMBL; AF058930; AAC14346.1; JOINED.
 CC EMBL; AF058931; AAC14346.1; JOINED.
 CC EMBL; AF058932; AAC14346.1; JOINED.
 CC EMBL; AF058933; AAC14346.1; JOINED.
 CC EMBL; AF058934; AAC14346.1; JOINED.
 CC EMBL; AF058935; AAC14346.1; JOINED.
 CC EMBL; AF058936; AAC14346.1; JOINED.
 CC EMBL; AF058937; AAC14346.1; JOINED.
 CC EMBL; AF058938; AAC14346.1; JOINED.
 CC EMBL; AF058939; AAC14346.1; JOINED.
 CC EMBL; AF058940; AAC14346.1; JOINED.
 CC EMBL; AF058941; AAC14346.1; JOINED.
 CC EMBL; AF058942; AAC14346.1; JOINED.
 CC EMBL; AF058943; AAC14346.1; JOINED.
 CC EMBL; AF058944; AAC14346.1; JOINED.
 CC EMBL; AF058945; AAC14346.1; JOINED.
 CC EMBL; AF058946; AAC14346.1; JOINED.
 CC EMBL; AF058947; AAC14346.1; JOINED.
 CC EMBL; AF058948; AAC14346.1; JOINED.
 CC EMBL; AF058949; AAC14346.1; JOINED.
 CC EMBL; AF058950; AAC14346.1; JOINED.
 CC EMBL; AF058951; AAC14346.1; JOINED.
 CC EMBL; AF058952; AAC14346.1; JOINED.
 CC EMBL; AF058953; AAC14346.1; JOINED.
 CC EMBL; AF058954; AAC14346.1; JOINED.
 CC EMBL; AF058955; AAC14346.1; JOINED.
 CC EMBL; AF058956; AAC14346.1; JOINED.
 CC EMBL; AF058957; AAC14346.1; JOINED.
 CC EMBL; AF058958; AAC14346.1; JOINED.
 CC EMBL; AF058959; AAC14346.1; JOINED.
 CC EMBL; AF058960; AAC14346.1; JOINED.
 CC EMBL; AF058961; AAC14346.1; JOINED.
 CC EMBL; AF058962; AAC14346.1; JOINED.
 CC EMBL; AF058963; AAC14346.1; JOINED.
 CC EMBL; AF058964; AAC14346.1; JOINED.
 CC EMBL; AF058965; AAC14346.1; JOINED.
 CC EMBL; AF058966; AAC14346.1; JOINED.
 CC EMBL; AF058967; AAC14346.1; JOINED.
 CC EMBL; AF058968; AAC14346.1; JOINED.
 CC EMBL; AF058969; AAC14346.1; JOINED.
 CC EMBL; AF058970; AAC14346.1; JOINED.
 CC EMBL; AF058971; AAC14346.1; JOINED.
 CC EMBL; AF058972; AAC14346.1; JOINED.
 CC EMBL; AF058973; AAC14346.1; JOINED.
 CC EMBL; AF058974; AAC14346.1; JOINED.
 CC EMBL; AF058975; AAC14346.1; JOINED.
 CC EMBL; AF058976; AAC14346.1; JOINED.
 CC EMBL; AF058977; AAC14346.1; JOINED.
 CC EMBL; AF058978; AAC14346.1; JOINED.
 CC EMBL; AF058979; AAC14346.1; JOINED.
 CC EMBL; AF058980; AAC14346.1; JOINED.
 CC EMBL; AF058981; AAC14346.1; JOINED.
 CC EMBL; AF058982; AAC14346.1; JOINED.
 CC EMBL; AF058983; AAC14346.1; JOINED.
 CC EMBL; AF058984; AAC14346.1; JOINED.
 CC EMBL; AF058985; AAC14346.1; JOINED.
 CC EMBL; AF058986; AAC14346.1; JOINED.
 CC EMBL; AF058987; AAC14346.1; JOINED.
 CC EMBL; AF058988; AAC14346.1; JOINED.
 CC EMBL; AF058989; AAC14346.1; JOINED.
 CC EMBL; AF058990; AAC14346.1; JOINED.
 CC EMBL; AF058991; AAC14346.1; JOINED.
 CC EMBL; AF058992; AAC14346.1; JOINED.
 CC EMBL; AF058993; AAC14346.1; JOINED.
 CC EMBL; AF058994; AAC14346.1; JOINED.
 CC EMBL; AF058995; AAC14346.1; JOINED.
 CC EMBL; AF058996; AAC14346.1; JOINED.
 CC EMBL; AF058997; AAC14346.1; JOINED.
 CC EMBL; AF058998; AAC14346.1; JOINED.
 CC EMBL; AF058999; AAC14346.1; JOINED.
 CC EMBL; AF059000; AAC14346.1; JOINED.
 CC EMBL; AF059001; AAC14346.1; JOINED.
 CC EMBL; AF059002; AAC14346.1; JOINED.
 CC EMBL; AF059003; AAC14346.1; JOINED.
 CC EMBL; AF059004; AAC14346.1; JOINED.
 CC EMBL; AF059005; AAC14346.1; JOINED.
 CC EMBL; AF059006; AAC14346.1; JOINED.
 CC EMBL; AF059007; AAC14346.1; JOINED.
 CC EMBL; AF059008; AAC14346.1; JOINED.
 CC EMBL; AF059009; AAC14346.1; JOINED.
 CC EMBL; AF059010; AAC14346.1; JOINED.
 CC EMBL; AF059011; AAC14346.1; JOINED.
 CC EMBL; AF059012; AAC14346.1; JOINED.
 CC EMBL; AF059013; AAC14346.1; JOINED.
 CC EMBL; AF059014; AAC14346.1; JOINED.
 CC EMBL; AF059015; AAC14346.1; JOINED.
 CC EMBL; AF059016; AAC14346.1; JOINED.
 CC EMBL; AF059017; AAC14346.1; JOINED.
 CC EMBL; AF059018; AAC14346.1; JOINED.
 CC EMBL; AF059019; AAC14346.1; JOINED.
 CC EMBL; AF059020; AAC14346.1; JOINED.
 CC EMBL; AF059021; AAC14346.1; JOINED.
 CC EMBL; AF059022; AAC14346.1; JOINED.
 CC EMBL; AF059023; AAC14346.1; JOINED.
 CC EMBL; AF059024; AAC14346.1; JOINED.
 CC EMBL; AF059025; AAC14346.1; JOINED.
 CC EMBL; AF059026; AAC14346.1; JOINED.
 CC EMBL; AF059027; AAC14346.1; JOINED.
 CC EMBL; AF059028; AAC14346.1; JOINED.
 CC EMBL; AF059029; AAC14346.1; JOINED.
 CC EMBL; AF059030; AAC14346.1; JOINED.
 CC EMBL; AF059031; AAC14346.1; JOINED.
 CC EMBL; AF059032; AAC14346.1; JOINED.
 CC EMBL; AF059033; AAC14346.1; JOINED.
 CC EMBL; AF059034; AAC14346.1; JOINED.
 CC EMBL; AF059035; AAC14346.1; JOINED.
 CC EMBL; AF059036; AAC14346.1; JOINED.
 CC EMBL; AF059037; AAC14346.1; JOINED.
 CC EMBL; AF059038; AAC14346.1; JOINED.
 CC EMBL; AF059039; AAC14346.1; JOINED.
 CC EMBL; AF059040; AAC14346.1; JOINED.
 CC EMBL; AF059041; AAC14346.1; JOINED.
 CC EMBL; AF059042; AAC14346.1; JOINED.
 CC EMBL; AF059043; AAC14346.1; JOINED.
 CC EMBL; AF059044; AAC14346.1; JOINED.
 CC EMBL; AF059045; AAC14346.1; JOINED.
 CC EMBL; AF059046; AAC14346.1; JOINED.
 CC EMBL; AF059047; AAC14346.1; JOINED.
 CC EMBL; AF059048; AAC14346.1; JOINED.
 CC EMBL; AF059049; AAC14346.1; JOINED.
 CC EMBL; AF059050; AAC14346.1; JOINED.
 CC EMBL; AF059051; AAC14346.1; JOINED.
 CC EMBL; AF059052; AAC14346.1; JOINED.
 CC EMBL; AF059053; AAC14346.1; JOINED.
 CC EMBL; AF059054; AAC14346.1; JOINED.
 CC EMBL; AF059055; AAC14346.1; JOINED.
 CC EMBL; AF059056; AAC14346.1; JOINED.
 CC EMBL; AF059057; AAC14346.1; JOINED.
 CC EMBL; AF059058; AAC14346.1; JOINED.
 CC EMBL; AF059059; AAC14346.1; JOINED.
 CC EMBL; AF059060; AAC14346.1; JOINED.
 CC EMBL; AF059061; AAC14346.1; JOINED.
 CC EMBL; AF059062; AAC14346.1; JOINED.
 CC EMBL; AF059063; AAC14346.1; JOINED.
 CC EMBL; AF059064; AAC14346.1; JOINED.
 CC EMBL; AF059065; AAC14346.1; JOINED.
 CC EMBL; AF059066; AAC14346.1; JOINED.
 CC EMBL; AF059067; AAC14346.1; JOINED.
 CC EMBL; AF059068; AAC14346.1; JOINED.
 CC EMBL; AF059069; AAC14346.1; JOINED.
 CC EMBL; AF059070; AAC14346.1; JOINED.
 CC EMBL; AF059071; AAC14346.1; JOINED.
 CC EMBL; AF059072; AAC14346.1; JOINED.
 CC EMBL; AF059073; AAC14346.1; JOINED.
 CC EMBL; AF059074; AAC14346.1; JOINED.
 CC EMBL; AF059075; AAC14346.1; JOINED.
 CC EMBL; AF059076; AAC14346.1; JOINED.
 CC EMBL; AF059077; AAC14346.1; JOINED.
 CC EMBL; AF059078; AAC14346.1; JOINED.
 CC EMBL; AF059079; AAC14346.1; JOINED.
 CC EMBL; AF059080; AAC14346.1; JOINED.
 CC EMBL; AF059081; AAC14346.1; JOINED.
 CC EMBL; AF059082; AAC14346.1; JOINED.
 CC EMBL; AF059083; AAC14346.1; JOINED.
 CC EMBL; AF059084; AAC14346.1; JOINED.
 CC EMBL; AF059085; AAC14346.1; JOINED.
 CC EMBL; AF059086; AAC14346.1; JOINED.
 CC EMBL; AF059087; AAC14346.1; JOINED.
 CC EMBL; AF059088; AAC14346.1; JOINED.
 CC EMBL; AF059089; AAC14346.1; JOINED.
 CC EMBL; AF059090; AAC14346.1; JOINED.
 CC EMBL; AF059091; AAC14346.1; JOINED.
 CC EMBL; AF059092; AAC14346.1; JOINED.
 CC EMBL; AF059093; AAC14346.1; JOINED.
 CC EMBL; AF059094; AAC14346.1; JOINED.
 CC EMBL; AF059095; AAC14346.1; JOINED.
 CC EMBL; AF059096; AAC14346.1; JOINED.
 CC EMBL; AF059097; AAC14346.1; JOINED.
 CC EMBL; AF059098; AAC14346.1; JOINED.
 CC EMBL; AF059099; AAC14346.1; JOINED.
 CC EMBL; AF059100; AAC14346.1; JOINED.
 CC EMBL; AF059101; AAC14346.1; JOINED.
 CC EMBL; AF059102; AAC14346.1; JOINED.
 CC EMBL; AF059103; AAC14346.1; JOINED.
 CC EMBL; AF059104; AAC14346.1; JOINED.
 CC EMBL; AF059105; AAC14346.1; JOINED.
 CC EMBL; AF059106; AAC14346.1; JOINED.
 CC EMBL; AF059107; AAC14346.1; JOINED.
 CC EMBL; AF059108; AAC14346.1; JOINED.
 CC EMBL; AF059109; AAC14346.1; JOINED.
 CC EMBL; AF059110; AAC14346.1; JOINED.
 CC EMBL; AF059111; AAC14346.1; JOINED.
 CC EMBL; AF059112; AAC14346.1; JOINED.
 CC EMBL; AF059113; AAC14346.1; JOINED.
 CC EMBL; AF059114; AAC14346.1; JOINED.
 CC EMBL; AF059115; AAC14346.1; JOINED.
 CC EMBL; AF059116; AAC14346.1; JOINED.
 CC EMBL; AF059117; AAC14346.1; JOINED.
 CC EMBL; AF059118; AAC14346.1; JOINED.
 CC EMBL; AF059119; AAC14346.1; JOINED.
 CC EMBL; AF059120; AAC14346.1; JOINED.
 CC EMBL; AF059121; AAC14346.1; JOINED.
 CC EMBL; AF059122; AAC14346.1; JOINED.
 CC EMBL; AF059123; AAC14346.1; JOINED.
 CC EMBL; AF059124; AAC14346.1; JOINED.
 CC EMBL; AF059125; AAC14346.1; JOINED.
 CC EMBL; AF059126; AAC14346.1; JOINED.
 CC EMBL; AF059127; AAC14346.1; JOINED.
 CC EMBL; AF059128; AAC14346.1; JOINED.
 CC EMBL; AF059129; AAC14346.1; JOINED.
 CC EMBL; AF059130; AAC14346.1; JOINED.
 CC EMBL; AF059131; AAC14346.1; JOINED.
 CC EMBL; AF059132; AAC14346.1; JOINED.
 CC EMBL; AF059133; AAC14346.1; JOINED.
 CC EMBL; AF059134; AAC14346.1; JOINED.
 CC EMBL; AF059135; AAC14346.1; JOINED.
 CC EMBL; AF059136; AAC14346.1; JOINED.
 CC EMBL; AF059137; AAC14346.1; JOINED.
 CC EMBL; AF059138; AAC14346.1; JOINED.
 CC EMBL; AF059139; AAC14346.1; JOINED.
 CC EMBL; AF059140; AAC14346.1; JOINED.
 CC EMBL; AF059141; AAC14346.1; JOINED.
 CC EMBL; AF059142; AAC14346.1; JOINED.
 CC EMBL; AF059143; AAC14346.1; JOINED.
 CC EMBL; AF059144; AAC14346.1; JOINED.
 CC EMBL; AF059145; AAC14346.1; JOINED.
 CC EMBL; AF059146; AAC14346.1; JOINED.
 CC EMBL; AF059147; AAC14346.1; JOINED.
 CC EMBL; AF059148; AAC14346.1; JOINED.
 CC EMBL; AF059149; AAC14346.1; JOINED.
 CC EMBL; AF059150; AAC14346.1; JOINED.
 CC EMBL; AF059151; AAC14346.1; JOINED.
 CC EMBL; AF059152; AAC14346.1; JOINED.
 CC EMBL; AF059153; AAC14346.1; JOINED.
 CC EMBL; AF059154; AAC14346.1; JOINED.
 CC EMBL; AF059155; AAC14346.1; JOINED.
 CC EMBL; AF059156; AAC14346.1; JOINED.
 CC EMBL; AF059157; AAC14346.1; JOINED.
 CC EMBL; AF059158; AAC14346.1; JOINED.
 CC EMBL; AF059159; AAC14346.1; JOINED.
 CC EMBL; AF059160; AAC14346.1; JOINED.
 CC EMBL; AF059161; AAC14346.1; JOINED.
 CC EMBL; AF059162; AAC14346.1; JOINED.
 CC EMBL; AF059163; AAC14346.1; JOINED.
 CC EMBL; AF059164; AAC14346.1; JOINED.
 CC EMBL; AF059165; AAC14346.1; JOINED.
 CC EMBL; AF059166; AAC14346.1; JOINED.
 CC EMBL; AF059167; AAC14346.1; JOINED.
 CC EMBL; AF059168; AAC14346.1; JOINED.
 CC EMBL; AF059169; AAC14346.1; JOINED.
 CC EMBL; AF059170; AAC14346.1; JOINED.
 CC EMBL; AF059171; AAC14346.1; JOINED.
 CC EMBL; AF059172; AAC14346.1; JOINED.
 CC EMBL; AF059173; AAC14346.1; JOINED.
 CC EMBL; AF059174; AAC14346.1; JOINED.
 CC EMBL; AF059175; AAC14346.1; JOINED.
 CC EMBL; AF059176; AAC14346.1; JOINED.
 CC EMBL; AF059177; AAC14346.1; JOINED.
 CC EMBL; AF059178; AAC14346.1; JOINED.
 CC EMBL; AF059179; AAC14346.1; JOINED.
 CC EMBL; AF059180; AAC14346.1; JOINED.
 CC EMBL; AF059181; AAC14346.1; JOINED.
 CC EMBL; AF059182; AAC14346.1; JOINED.
 CC EMBL; AF059183; AAC14346.1; JOINED.
 CC EMBL; AF059184; AAC14346.1; JOINED.
 CC EMBL; AF059185; AAC14346.1; JOINED.
 CC EMBL; AF059186; AAC14346.1; JOINED.
 CC EMBL; AF059187; AAC14346.1; JOINED.
 CC EMBL; AF059188; AAC14346.1; JOINED.
 CC EMBL; AF059189; AAC14346.1; JOINED.
 CC EMBL; AF059190; AAC14346.1; JOINED.
 CC EMBL; AF059191; AAC14346.1; JOINED.
 CC EMBL; AF059192; AAC14346.1; JOINED.
 CC EMBL; AF059193; AAC14346.1; JOINED.
 CC EMBL; AF059194; AAC14346.1; JOINED.
 CC EMBL; AF059195; AAC14346.1; JOINED.
 CC EMBL; AF059196; AAC14346.1; JOINED.
 CC EMBL; AF059197; AAC14346.1; JOINED.
 CC EMBL; AF059198; AAC14346.1; JOINED.
 CC EMBL; AF059199; AAC14346.1; JOINED.
 CC EMBL; AF059200; AAC14346.1; JOINED.
 CC EMBL; AF059201; AAC14346.1; JOINED.
 CC EMBL; AF059202; AAC14346.1; JOINED.
 CC EMBL; AF059203; AAC14346.1; JOINED.
 CC EMBL; AF059204; AAC14346.1; JOINED.
 CC EMBL; AF059205; AAC14346.1; JOINED.
 CC EMBL; AF059206; AAC14346.1; JOINED.
 CC EMBL; AF059207; AAC14346.1; JOINED.
 CC EMBL; AF059208; AAC14346.1; JOINED.
 CC EMBL; AF059209; AAC14346.1; JOINED.
 CC EMBL; AF059210; AAC14346.1; JOINED.
 CC EMBL; AF059211; AAC14346.1; JOINED.
 CC EMBL; AF059212; AAC14346.1; JOINED.
 CC EMBL; AF059213; AAC14346.1; JOINED.
 CC EMBL; AF059214; AAC14346.1; JOINED.
 CC EMBL; AF059215; AAC14346.1; JOINED.
 CC EMBL; AF059216; AAC14346.1; JOINED.
 CC EMBL; AF059217; AAC14346.1; JOINED.
 CC EMBL; AF059218; AAC14346.1; JOINED.
 CC EMBL; AF059219; AAC14346.1; JOINED.
 CC EMBL; AF059220; AAC14346.1; JOINED.
 CC EMBL; AF059221; AAC14346.1; JOINED.
 CC EMBL; AF059222; AAC14346.1; JOINED.
 CC EMBL; AF059223; AAC14346.1; JOINED.
 CC EMBL; AF059224; AAC14346.1; JOINED.
 CC EMBL; AF059225; AAC14346.1; JOINED.
 CC EMBL; AF059226; AAC14346.1; JOINED.
 CC EMBL; AF059227; AAC14346.1; JOINED.
 CC EMBL; AF059228; AAC14346.1; JOINED.
 CC EMBL; AF059229; AAC14346.1; JOINED.
 CC EMBL; AF059230; AAC14346.1; JOINED.
 CC EMBL; AF059231; AAC14346.1; JOINED.
 CC EMBL; AF059232; AAC14346.1; JOINED


```
CC -----
DR EMBL; M33874; AAB02039.1; -.
DR HSSP; P00740; 1EDM.
DR InterPro; IPR002110; ANK.
DR InterPro; IPR000152; Asx hydroxyl_S.
DR InterPro; IPR000742; EGF 2.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR001438; EGF_11.
DR InterPro; IPR006209; EGF_11e.
DR InterPro; IPR002049; Laminin_EGF.
DR InterPro; IPR008297; Notch.
DR InterPro; IPR000800; Notch_dom.
DR Pfam; PF00023; ank; 6.
DR Pfam; PF00008; EGF; 36.
DR Pfam; PF00066; notch; 3.
DR PIRSF; PIRSF002279; Notch; 1.
DR PRINTS; PR00010; EGFLAMOD.
DR PRINTS; PR00011; EGFLAMININ.
DR PRINTS; PR01452; NOTCH.
DR SMART; SM00248; ANK; 6.
DR SMART; SM00179; EGF_CA; 24.
DR SMART; SM00004; NL; 2.
DR PROSITE; PS50297; ANK_REPEAT; 1.
DR PROSITE; PS50089; ANK_REPEAT; 4.
DR PROSITE; PS00010; ASX_HYDROXYL; 23.
DR PROSITE; PS00022; EGF_1; 34.
DR PROSITE; PS01186; EGF_2; 29.
DR PROSITE; PS50026; EGF_3; 36.
DR PROSITE; PS01187; EGF_CA; 21.
KW Differentiation; Neurogenesis; Repeat; ANK repeat; EGF-like domain;
KW Transmembrane; Signal; Glycoprotein.
FT SIGNAL 1 19
FT CHAIN 20 2324
FT DOMAIN 20 1728
FT TRANSMM 1729 1750
FT DOMAIN 1751 2524
FT DOMAIN 20 57
FT DOMAIN 58 99
FT DOMAIN 102 140
FT DOMAIN 141 177
FT DOMAIN 179 215
FT DOMAIN 217 254
FT DOMAIN 256 292
FT DOMAIN 294 332
FT DOMAIN 334 370
FT DOMAIN 371 409
FT DOMAIN 411 449
FT DOMAIN 451 487
FT DOMAIN 489 525
FT DOMAIN 527 563
FT DOMAIN 565 600
FT DOMAIN 602 638
FT DOMAIN 640 675
FT DOMAIN 677 713
FT DOMAIN 715 750
FT DOMAIN 752 788
FT DOMAIN 790 826
FT DOMAIN 828 866
FT DOMAIN 868 904
FT DOMAIN 906 942
FT DOMAIN 944 980
FT DOMAIN 982 1018
FT DOMAIN 1020 1056
FT DOMAIN 1058 1094
FT DOMAIN 1096 1140
FT DOMAIN 1144 1180
FT DOMAIN 1182 1218
FT DOMAIN 1220 1264
FT DOMAIN 1266 1304
FT DOMAIN 1306 1346
FT DOMAIN 1347 1383
FT DOMAIN 1386 1424
FT REPEAT 1441 1478
-----
FT REPEAT 1479 1520
FT REPEAT 1521 1560
FT REPEAT 1561 1600
FT REPEAT 1601 1640
FT REPEAT 1641 1680
FT REPEAT 1681 1720
FT REPEAT 1721 1760
FT REPEAT 1761 1800
FT REPEAT 1801 1840
FT REPEAT 1841 1880
FT REPEAT 1881 1920
FT REPEAT 1921 1960
FT REPEAT 1961 2000
FT REPEAT 2001 2040
FT REPEAT 2041 2080
FT REPEAT 2081 2120
FT REPEAT 2121 2160
FT REPEAT 2161 2200
FT REPEAT 2201 2240
FT REPEAT 2241 2280
FT REPEAT 2281 2320
FT REPEAT 2321 2360
FT REPEAT 2361 2400
FT REPEAT 2401 2440
FT REPEAT 2441 2480
FT REPEAT 2481 2520
FT REPEAT 2521 2560
FT REPEAT 2561 2600
FT REPEAT 2601 2640
FT REPEAT 2641 2680
FT REPEAT 2681 2720
FT REPEAT 2721 2760
FT REPEAT 2761 2800
FT REPEAT 2801 2840
FT REPEAT 2841 2880
FT REPEAT 2881 2920
FT REPEAT 2921 2960
FT REPEAT 2961 3000
FT REPEAT 3001 3040
FT REPEAT 3041 3080
FT REPEAT 3081 3120
FT REPEAT 3121 3160
FT REPEAT 3161 3200
FT REPEAT 3201 3240
FT REPEAT 3241 3280
FT REPEAT 3281 3320
FT REPEAT 3321 3360
FT REPEAT 3361 3400
FT REPEAT 3401 3440
FT REPEAT 3441 3480
FT REPEAT 3481 3520
FT REPEAT 3521 3560
FT REPEAT 3561 3600
FT REPEAT 3601 3640
FT REPEAT 3641 3680
FT REPEAT 3681 3720
FT REPEAT 3721 3760
FT REPEAT 3761 3800
FT REPEAT 3801 3840
FT REPEAT 3841 3880
FT REPEAT 3881 3920
FT REPEAT 3921 3960
FT REPEAT 3961 4000
FT REPEAT 4001 4040
FT REPEAT 4041 4080
FT REPEAT 4081 4120
FT REPEAT 4121 4160
FT REPEAT 4161 4200
FT REPEAT 4201 4240
FT REPEAT 4241 4280
FT REPEAT 4281 4320
FT REPEAT 4321 4360
FT REPEAT 4361 4400
FT REPEAT 4401 4440
FT REPEAT 4441 4480
FT REPEAT 4481 4520
FT REPEAT 4521 4560
FT REPEAT 4561 4600
FT REPEAT 4601 4640
FT REPEAT 4641 4680
FT REPEAT 4681 4720
FT REPEAT 4721 4760
FT REPEAT 4761 4800
FT REPEAT 4801 4840
FT REPEAT 4841 4880
FT REPEAT 4881 4920
FT REPEAT 4921 4960
FT REPEAT 4961 5000
FT REPEAT 5001 5040
FT REPEAT 5041 5080
FT REPEAT 5081 5120
FT REPEAT 5121 5160
FT REPEAT 5161 5200
FT REPEAT 5201 5240
FT REPEAT 5241 5280
FT REPEAT 5281 5320
FT REPEAT 5321 5360
FT REPEAT 5361 5400
FT REPEAT 5401 5440
FT REPEAT 5441 5480
FT REPEAT 5481 5520
FT REPEAT 5521 5560
FT REPEAT 5561 5600
FT REPEAT 5601 5640
FT REPEAT 5641 5680
FT REPEAT 5681 5720
FT REPEAT 5721 5760
FT REPEAT 5761 5800
FT REPEAT 5801 5840
FT REPEAT 5841 5880
FT REPEAT 5881 5920
FT REPEAT 5921 5960
FT REPEAT 5961 6000
FT REPEAT 6001 6040
FT REPEAT 6041 6080
FT REPEAT 6081 6120
FT REPEAT 6121 6160
FT REPEAT 6161 6200
FT REPEAT 6201 6240
FT REPEAT 6241 6280
FT REPEAT 6281 6320
FT REPEAT 6321 6360
FT REPEAT 6361 6400
FT REPEAT 6401 6440
FT REPEAT 6441 6480
FT REPEAT 6481 6520
FT REPEAT 6521 6560
FT REPEAT 6561 6600
FT REPEAT 6601 6640
FT REPEAT 6641 6680
FT REPEAT 6681 6720
FT REPEAT 6721 6760
FT REPEAT 6761 6800
FT REPEAT 6801 6840
FT REPEAT 6841 6880
FT REPEAT 6881 6920
FT REPEAT 6921 6960
FT REPEAT 6961 7000
FT REPEAT 7001 7040
FT REPEAT 7041 7080
FT REPEAT 7081 7120
FT REPEAT 7121 7160
FT REPEAT 7161 7200
FT REPEAT 7201 7240
FT REPEAT 7241 7280
FT REPEAT 7281 7320
FT REPEAT 7321 7360
FT REPEAT 7361 7400
FT REPEAT 7401 7440
FT REPEAT 7441 7480
FT REPEAT 7481 7520
FT REPEAT 7521 7560
FT REPEAT 7561 7600
FT REPEAT 7601 7640
FT REPEAT 7641 7680
FT REPEAT 7681 7720
FT REPEAT 7721 7760
FT REPEAT 7761 7800
FT REPEAT 7801 7840
FT REPEAT 7841 7880
FT REPEAT 7881 7920
FT REPEAT 7921 7960
FT REPEAT 7961 8000
FT REPEAT 8001 8040
FT REPEAT 8041 8080
FT REPEAT 8081 8120
FT REPEAT 8121 8160
FT REPEAT 8161 8200
FT REPEAT 8201 8240
FT REPEAT 8241 8280
FT REPEAT 8281 8320
FT REPEAT 8321 8360
FT REPEAT 8361 8400
FT REPEAT 8401 8440
FT REPEAT 8441 8480
FT REPEAT 8481 8520
FT REPEAT 8521 8560
FT REPEAT 8561 8600
FT REPEAT 8601 8640
FT REPEAT 8641 8680
FT REPEAT 8681 8720
FT REPEAT 8721 8760
FT REPEAT 8761 8800
FT REPEAT 8801 8840
FT REPEAT 8841 8880
FT REPEAT 8881 8920
FT REPEAT 8921 8960
FT REPEAT 8961 9000
FT REPEAT 9001 9040
FT REPEAT 9041 9080
FT REPEAT 9081 9120
FT REPEAT 9121 9160
FT REPEAT 9161 9200
FT REPEAT 9201 9240
FT REPEAT 9241 9280
FT REPEAT 9281 9320
FT REPEAT 9321 9360
FT REPEAT 9361 9400
FT REPEAT 9401 9440
FT REPEAT 9441 9480
FT REPEAT 9481 9520
FT REPEAT 9521 9560
FT REPEAT 9561 9600
FT REPEAT 9601 9640
FT REPEAT 9641 9680
FT REPEAT 9681 9720
FT REPEAT 9721 9760
FT REPEAT 9761 9800
FT REPEAT 9801 9840
FT REPEAT 9841 9880
FT REPEAT 9881 9920
FT REPEAT 9921 9960
FT REPEAT 9961 10000
```

FT	DISULFID	856	865	BY SIMILARITY.
FT	DISULFID	872	883	BY SIMILARITY.
FT	DISULFID	877	892	BY SIMILARITY.
FT	DISULFID	894	903	BY SIMILARITY.
FT	DISULFID	910	921	BY SIMILARITY.
FT	DISULFID	915	930	BY SIMILARITY.
FT	DISULFID	932	941	BY SIMILARITY.
FT	DISULFID	986	997	BY SIMILARITY.
FT	DISULFID	991	1006	BY SIMILARITY.
FT	DISULFID	1008	1017	BY SIMILARITY.
FT	DISULFID	1024	1035	BY SIMILARITY.
FT	DISULFID	1029	1044	BY SIMILARITY.
FT	DISULFID	1046	1054	BY SIMILARITY.
FT	DISULFID	1062	1073	BY SIMILARITY.
FT	DISULFID	1067	1082	BY SIMILARITY.
FT	DISULFID	1084	1093	BY SIMILARITY.
FT	DISULFID	1100	1121	BY SIMILARITY.
FT	DISULFID	1115	1130	BY SIMILARITY.

Query Match	Score 80;	DB 1;	Length 2524;
Best Local Similarity	57.64;		
Best Local Similarity	41.08;		
Best Local Similarity	41.08;		

Dy 1 CDXXXCXXKXGNGXCDDXXCNNAACXXDXDC 31
Dd 1447 CENFQCSLETADNKVCNANONNHACGDGGDC 1477

RESULT 14	STANDARD	PRT	1295 AA.
GLP1_CAEEL			
AC	P13508;		
DT	01-JUN-1990 (Rel. 13, Created)		
DT	01-JAN-1990 (Rel. 13, Last sequence update)		
DT	15-MAR-2004 (Rel. 43, Last annotation update)		
DE	GLP-1 protein precursor.		
GN	GLP-1 OR EMB-33 OR F02A9.6.		
OS	Caenorhabditis elegans.		
OC	Eukaryotes; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae;		
OC	Rhabditidae; Pelodierinae; Caenorhabditis.		
OX	NCBI_TaxId=6239;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=Bristol N2;		
RC	MEDLINE=89363787; PubMed=2758466;		
RA	Yocham J., Greenwald I.;		
RT	"gfp-1 and lin-12, genes implicated in distinct cell-cell		
RT	interactions in C. elegans, encode similar transmembrane proteins.",		
RT	Cell 58:553-563 (1989).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=Bristol N2;		
RC	MEDLINE=94150718; PubMed=7906398;		
RA	Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,		
RA	Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.		
RA	Crauxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,		
RA	Pollon L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,		
RA	Johnson L., Jones M., Kerzshaw J., Kitzien J., Laister N.,		
RA	Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,		
RA	Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkhen R.		
RA	Sims M., Smaldon N., Smith A., Smith M., Sonhammer E., Staden R.,		
RA	Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,		
RA	Waterston R., Watson A., Weinstein L., Wilkinson-Sprat J.,		
RA	Whitman P.;		
RT	"2.2 Mb of contiguous nucleotide sequence from chromosome III of C.		
RT	elegans.";		
RT	Nature 368:32-38 (1994).		
RN	[3]		
RP	DELETION OF 1174-1295.		
RC	MEDLINE=9151288; PubMed=188136;		
RA	Mango S.E., Maine E.M., Kimble J.;		
RT	"Carboxy-terminal truncation activates glp-1 protein to specify		
RT	valvul fate in Caenorhabditis elegans.";		
RT	VT		

RL	Nature 352:811-815 (1991).	
RN	(4)	
RN	CHARACTERIZATION OF FUNCTION OF THE ANK-REPEATS.	
RX	MEDLINE=93354444; PubMed=8350921;	
RA	Roehl H., Kimble J.;	
RT	"Control of cell fate in C. elegans by a GLP-1 peptide consisting	
RL	primarily of ankyrin repeats.";	
RL	Nature 364:632-635 (1993).	
RP	(5)	
RP	FUNCTION.	
RX	MEDLINE=94208066; PubMed=8156602;	
RA	Mello C.C., Draper B.W., Priess J.R.;	
RT	"The maternal genes apx-1 and glp-1 and establishment of	
RL	dorsal-ventral polarity in the early C. elegans embryo.";	
RL	Cell 77:95-106 (1994).	
CC	-1- FUNCTION: Involved in the specification of the cell fates of the	
CC	blastomeres, ABA and APA. Proper signaling by glp-1 induces ABA	
CC	descendants to produce anterior pharyngeal cells, and ABA	
CC	descendants to adopt a different fate. Contributes to the	
CC	establishment of the dorsal-ventral axis in early embryos.	
CC	-1- SUBCELLULAR LOCATION: Type I membrane protein.	
CC	-1- DEVELOPMENTAL STAGE: Acts on ABP development during 4-cell and	
CC	12-cell stages, and on ABA development during 12-cell and 28-cell	
CC	stages.	
CC	-1- SIMILARITY: HIGH, TO C.ELEGANS LIN-12.	
CC	-1- SIMILARITY: Contains 10 EGF-like domains.	
CC	-1- SIMILARITY: Contains 3 Lin/Notch repeats.	
CC	-1- SIMILARITY: Contains 5 ANK repeats.	
CC	-----	
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration	
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -	
CC	the European Bioinformatics Institute. There are no restrictions on its	
CC	use by non-profit institutions as long as its content is in no way	
CC	modified and this statement is not removed. Usage by and for commercial	
CC	entities requires a license agreement (See http://www.isb.ch/announce/isb-sib.ch).	
CC	or send an email to license@isb-sib.ch .	
CC	-----	
CC	EMBL; M25580; AAA28058.1; -;	
DR	EMBL; Z19555; CAA79620.1; -;	
DR	EMBL; Z29116; CAA79620.1; JOINED.	
DR	EMBL; Z29116; CAA82373.1; -;	
DR	EMBL; Z19555; CAA82373.1; JOINED.	
DR	PIR; A32901; A32901.	
DR	HSSP; P00740; IEDM.	
DR	GermOnline; 193793; -;	
DR	WormPep; F02A3.6; CE00237.	
DR	InterPro; IPR002110; ANK.	
DR	InterPro; IPR000152; Asx_hydroxyl_S.	
DR	InterPro; IPR000742; EGF 2.	
DR	InterPro; IPR001881; EGF CA.	
DR	InterPro; IPR006209; EGF_like.	
DR	InterPro; IPR000800; Notch_dom.	
DR	Pfam; PF00023; ank; 5	
DR	Pfam; PF00008; EGF; 10.	
DR	Pfam; PF00066; notch; 3.	
DR	PRINTS; PR01452; NOTCH.	
DR	SMART; SMO0248; ANK; 5.	
DR	SMART; SMO0179; EGF CA; 1.	
DR	SMART; SMO0004; NL; 3.	
DR	PROSITE; PSS0297; ANK_REPEAT, 1.	
DR	PROSITE; PSS0088; ANK_REPEAT, 3.	
DR	PROSITE; PS00010; ASX_HYDROXYL; 2.	
DR	PROSITE; PS00022; EGF_1; 10.	
DR	PROSITE; PS01186; EGF_2; 8.	
DR	PROSITE; PSS0026; EGF_3; 10.	
DR	PROSITE; PS01187; EGF CA; 1.	
KW	Differentiation; Repeat; ANK repeat; EGF-like domain; Transmembrane;	
KM	Glycoprotein; Signal.	
FT	SIGNAL	15
FT	CHAIN	16 1295
FT	DOMAIN	16 764
FT	TRANSMEM	765 786
FT	DOMAIN	787 1295
FT	CYTOPLASMIC (POTENTIAL).	
FT	POTENTIAL.	
FT	GLP-1 PROTEIN (POTENTIAL).	
FT	EXTRACELLULAR (POTENTIAL).	
FT	POTENTIAL.	
FT	CYTOPLASMIC (POTENTIAL).	
FT	POTENTIAL.	

FT DOMAIN 19 58 EGF-LIKE 1.
 FT DOMAIN 117 152 EGF-LIKE 2.
 FT DOMAIN 154 190 EGF-LIKE 3.
 FT DOMAIN 190 230 EGF-LIKE 4.
 FT DOMAIN 232 269 EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 272 308 EGF-LIKE 6.
 FT DOMAIN 316 359 EGF-LIKE 7.
 FT DOMAIN 369 406 EGF-LIKE 8.
 FT DOMAIN 443 479 EGF-LIKE 9.
 FT REPEAT 493 527 EGF-LIKE 10.
 FT REPEAT 527 568 LTN/NOTCH 1.
 FT REPEAT 568 608 LTN/NOTCH 2.
 FT REPEAT 608 688 LTN/NOTCH 3.
 FT REPEAT 688 790 ANK 1.
 FT REPEAT 790 990 ANK 2.
 FT REPEAT 994 1023 ANK 3.
 FT REPEAT 1030 1062 ANK 4.
 FT REPEAT 1074 1103 ANK 5.
 FT REPEAT 1107 1136 BY SIMILARITY.
 FT DISULFID 23 35 BY SIMILARITY.
 FT DISULFID 29 46 BY SIMILARITY.
 FT DISULFID 48 57 BY SIMILARITY.
 FT DISULFID 121 131 BY SIMILARITY.
 FT DISULFID 126 140 BY SIMILARITY.
 FT DISULFID 142 151 BY SIMILARITY.
 FT DISULFID 158 169 BY SIMILARITY.
 FT DISULFID 163 178 BY SIMILARITY.
 FT DISULFID 180 189 BY SIMILARITY.
 FT DISULFID 194 206 BY SIMILARITY.
 FT DISULFID 201 218 BY SIMILARITY.
 FT DISULFID 220 229 BY SIMILARITY.
 FT DISULFID 236 248 BY SIMILARITY.
 FT DISULFID 242 257 BY SIMILARITY.
 FT DISULFID 259 268 BY SIMILARITY.
 FT DISULFID 275 286 BY SIMILARITY.
 FT DISULFID 280 296 BY SIMILARITY.
 FT DISULFID 298 307 BY SIMILARITY.
 FT DISULFID 373 384 BY SIMILARITY.
 FT DISULFID 378 394 BY SIMILARITY.
 FT DISULFID 396 405 BY SIMILARITY.
 FT DISULFID 411 422 BY SIMILARITY.
 FT DISULFID 416 431 BY SIMILARITY.
 FT DISULFID 433 442 BY SIMILARITY.
 FT DISULFID 450 461 BY SIMILARITY.
 FT DISULFID 469 478 BY SIMILARITY.
 FT DISULFID 485 497 BY SIMILARITY.
 FT CARBOHYD 244 244 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 245 245 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 333 333 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 381 381 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 609 609 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 675 675 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 1295 AA; 144078 MW; 422AADDA2DEEF3B4 CRC64;

Query Match 56.8%; Score 79; DB 1; Length 1235;
 Best Local Similarity 46.2%; Pred. No. 0.0024;
 Matches 12; Conservative 2; Mismatches 12; Indels 0; Gaps 0;

Oy 6 CXXXXXGXCXDXCNNAACXXDXGXC 31
 Db 542 CADQFANGVCNQECNNECLYDGLDC 567

RESULT 15
 NC4 MOUSE STANDARD: PRT; 1964 AA.
 AC P31695.035442; O88314; O62389; Q62390; Q9R1W9; Q9R1X0;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Neurogenic locus notch homolog protein 4 precursor (Notch 4)
 GN [Contains: Transforming protein Int-3].
 OS NOTCH4 OR INT3 OR INT-3.
 Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OK NCBI_Taxid:10090;
 RN (1)
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92194507; PubMed=1312643;
 RA Robbins J., Blondel B.J., Gallahan D., Callahan R.;
 RT "Mouse mammary tumor gene int-3: a member of the notch gene family
 RT transforms mammary epithelial cells."; J. Virol. 66:2594-2599(1992).
 RN (2)
 RP REVISIONS, SEQUENCE FROM N.A.
 RX MEDLINE=97294599; PubMed=9150355;
 RA Gallahan D., Callahan R.;
 RT "The mouse mammary tumor associated gene INT3 is a unique member of
 RT the NOTCH gene family (NOTCH4)."; Oncogene 14:1883-1890(1997).
 RN (3)
 RP SEQUENCE FROM N.A.
 RX TISSUE=Lung, and Testis;
 RX MEDLINE=96281668; PubMed=8681805;
 RA Uytendaele H., Marazzi G., Wu G., Yan Q., Sasseon D., Kitajewski J.;
 RT "Notch4/int-3, a mammary proto-oncogene, is an endothelial
 RT cell-specific mammalian Notch gene."; Development 122:2251-2259(1996).
 RN (4)
 RP SEQUENCE FROM N.A.
 RA Rowen L., Mahairs G., Qin S., Ahern M.E., Dankers C., Lasky S.,
 RA Loretz C., Schmidt S., Tipson S., Traicoff R., Zackroff K., Hood L.;
 RT "Sequence of the mouse major histocompatibility locus class III
 RT region."; Submitted (OCT-1997) to the EMBL/Genbank/DBJ databases.
 RN (5)
 RP SEQUENCE OF 1436-1600 FROM N.A.
 RX MEDLINE=99252212; PubMed=10233982;
 RA Lee J.-S., Haruna T., Ishimoto A., Honjo T., Yanagawa S.-I.;
 RT "Intracisternal type A particle-mediated activation of the Notch4/int3
 RT gene in a mouse mammary tumor: generation of truncated Notch4/int3
 RT mRNAs by retroviral splicing events."; J. Virol. 73:5166-5171(1999).
 RN (6)
 RP FUNCTION.
 RX MEDLINE=21244657; PubMed=11344305;
 RA Uytendaele H., Ho J., Rossant J., Kitajewski J.;
 RT "Vascular patterning defects associated with expression of activated
 RT Notch4 in embryonic endothelium."; Proc. Natl. Acad. Sci. U.S.A. 98:5643-5648(2001).
 RN (7)
 RP SEQUENCE OF 1463-1964, POST-TRANSLATIONAL PROCESSING, AND MUTAGENESIS
 RP OF VAL-1463.
 RX MEDLINE=21523956; PubMed=11518718;
 RA Saxena M.T., Schroeter E.H., Mumm J.S., Kopan R.;
 RT "Murine notch homologs (N1-4) undergo presenilin-dependent
 RT proteolysis."; J. Biol. Chem. 276:40268-40273(2001).
 RN (8)
 RP POST-TRANSLATIONAL PROCESSING.
 RX MEDLINE=21374376; PubMed=11459941;
 RA Mitani T., Taniguchi Y., Aoki T., Hashimoto N., Honjo T.;
 RT "Conservation of the biochemical mechanisms of signal transduction
 RT among mammalian Notch family members."; Proc. Natl. Acad. Sci. U.S.A. 98:9026-9031(2001).
 CC -1- FUNCTION: Functions as a receptor for membrane-bound ligands
 CC Jagged1, Jagged2 and Delta to regulate cell-fate determination.
 CC upon ligand activation through the released notch intracellular
 CC domain (NICD) it forms a transcriptional activator complex with
 CC RBP-J kappa and activates genes of the enhancer of split locus.
 CC Affects the implementation of differentiation, proliferation and
 CC apoptotic programs (By similarity). May regulate branching
 CC morphogenesis in the developing vascular system.
 CC -1- SUBUNIT: Heterodimer of a C-terminal fragment N(TM) and a N-
 CC terminal fragment N(EC) which are probably linked by disulfide
 CC bonds.

CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Following
 CC proteolytical processing NICD is translocated to the nucleus.
 CC -1- TISSUE SPECIFICITY: Highly expressed in lung, moderately in heart
 CC kidney, and at lower levels in the ovary and skeletal muscle. A
 CC very low expression is seen in the brain, intestine, liver and
 CC testis.
 CC -1- DEVELOPMENTAL STAGE: Highly expressed in endothelial cells during
 CC embryonic development from 9.0 dpc.
 CC -1- PTM: Synthesized in the endoplasmic reticulum as an inactive form
 CC which is proteolytically cleaved by a furin-like convertase in the
 CC trans-Golgi network before it reaches the plasma membrane to yield
 CC an active, ligand-accessible form. Cleavage results in a C-
 CC terminal fragment N(TW) and a N-terminal fragment N(EC). Following
 CC ligand binding, it is cleaved by TNF-alpha converting enzyme
 CC (TACE) to yield a membrane-associated intermediate fragment called
 CC notch extracellular truncation (NEXT). This fragment is then
 CC cleaved by presenilin dependent gamma-secretase to release a
 CC notch-derived peptide containing the intracellular domain (NICD)
 CC from the membrane.
 CC -1- PTM: Phosphorylated.
 CC -1- DISEASE: Loss of the extracellular domain causes constitutive
 CC activation of the Notch protein, which leads to hyperproliferation
 CC of glandular epithelial tissues and development of mammary
 CC carcinomas.
 CC -1- SIMILARITY: Belongs to the NOTCH family.
 CC -1- SIMILARITY: Contains 29 EGF-like domains.
 CC -1- SIMILARITY: Contains 3 lin/Notch repeats.
 CC -1- SIMILARITY: Contains 5 ANK repeats.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M80456; AAB38377.1; -;
 DR EMBL; U43691; AAC52630.1; -;
 DR EMBL; U43691; AAC52631.1; -;
 DR EMBL; AF030001; AAB82004.1; -;
 DR EMBL; AB016771; BAA32281.1; ALT SEQ.
 DR EMBL; AB016772; BAA32283.1; ALT INIT.
 DR EMBL; AB016773; BAA32284.1; ALT INIT.
 DR EMBL; AB016774; BAA32285.1; -;
 DR PIR; A38072; TVWV73.
 DR PIR; T09059; T09059.
 DR HSSP; P08709; 1BP9.
 DR MGD; MG1107471; Notch4.
 DR InterPro; IPR002110; ANK.
 DR InterPro; IPR000152; Asx_hydroxyl_S.
 DR InterPro; IPR000742; EGF_2.
 DR InterPro; IPR01881; EGF_CA.
 DR InterPro; IPR001438; EGF_II.
 DR InterPro; IPR006209; EGF_like.
 DR InterPro; IPR002049; laminin_EGF.
 DR InterPro; IPR008297; Notch.
 DR InterPro; IPR008000; Notch_dom.
 DR Pfam; PF00023; ank; 6.
 DR Pfam; PF00008; EGF; 27.
 DR Pfam; PF00066; notch; 2.
 DR PIRSF; PIRSF002279; Notch; 1.
 DR PRINTS; PRO0010; EGFBL00D.
 DR PRINTS; PRO0011; EGFBLAMIN.
 DR PRINTS; PRO1452; NOTCH.
 DR SMART; SM00248; ANK; 6.
 DR SMART; SM00179; EGF_CA; 11.
 DR SMART; SM00004; NL; 2.
 DR PROSITE; PS50297; ANK_REPEAT; 1.
 DR PROSITE; PS50088; ANK_REPEAT; 5.
 DR PROSITE; PS00010; ASK_HYDROXYL; 11.
 DR PROSITE; PS00022; EGF_1; 28.
 DR PROSITE; PS01186; EGF_2; 21.

DR PROSITE; PS50026; EGF_3; 27.
 DR PROSITE; PS01187; EGF_CA; 9.
 KW Receptor; Transcription regulation; Activator; Differentiation;
 KW Developmental protein; Repeat; ANK repeat; EGF-like domain;
 KW Transmembrane; Glycoprotein; Signal; Phosphorylation; Proto-oncogene.
 FT SIGNAL 1 20
 FT CHAIN 21 1964
 FT CHAIN 1411 1964
 FT CHAIN 1428 1964
 FT CHAIN 1463 1964
 FT DOMAIN 21 1443
 FT TRANSMEM 1444 1464
 FT DOMAIN 1465 1964
 FT DOMAIN 21 1964
 FT DOMAIN 61 112
 FT DOMAIN 115 152
 FT DOMAIN 153 189
 FT DOMAIN 191 229
 FT DOMAIN 231 271
 FT DOMAIN 273 309
 FT DOMAIN 311 350
 FT DOMAIN 352 388
 FT DOMAIN 389 427
 FT DOMAIN 429 470
 FT DOMAIN 472 508
 FT DOMAIN 510 546
 FT DOMAIN 548 584
 FT DOMAIN 586 622
 FT DOMAIN 623 656
 FT DOMAIN 658 686
 FT DOMAIN 688 724
 FT DOMAIN 726 762
 FT DOMAIN 764 800
 FT DOMAIN 803 839
 FT DOMAIN 841 877
 FT DOMAIN 878 924
 FT DOMAIN 926 962
 FT DOMAIN 964 1000
 FT DOMAIN 1002 1040
 FT DOMAIN 1042 1081
 FT DOMAIN 1083 1122
 FT DOMAIN 1126 1167
 FT REPEAT 1168 1208
 FT REPEAT 1209 1242
 FT REPEAT 1243 1282
 FT DOMAIN 926 962 EGF-LIKE 23.
 FT DOMAIN 964 1000 EGF-LIKE 24.
 FT DOMAIN 1002 1040 EGF-LIKE 25.
 FT DOMAIN 1042 1081 EGF-LIKE 26.
 FT DOMAIN 1083 1122 EGF-LIKE 27.
 FT DOMAIN 1126 1167 EGF-LIKE 28.
 FT REPEAT 1168 1208 EGF-LIKE 29.
 FT REPEAT 1209 1242 LIN/NOTCH 1.
 FT REPEAT 1243 1282 LIN/NOTCH 2.
 FT REPEAT 1243 1282 LIN/NOTCH 3.

Query Match 56.8% Score 79; DB 1; Length 1964;
 Best Local Similarity 50.0%; Pred. No. 0.00036;
 Matches 13; Conservative 1; Mismatches 12; Indels 0; Gaps 0;
 Qy 6 CXXXXXNGXCDXXCNNAACXXDXGDC 31
 Db 1255 CRDHFHNGHCXKGNNAACGMDGDC 1280

Search completed: May 19, 2004, 15:58:21
 Job time : 0.785805 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using SW model

Run on: May 19, 2004, 15:54:24 ; Search time 2.98606 seconds
(without alignments)
3275.577 Million cell updates/sec

Title: US-09-903-199-4

Sequence: 1 CDXXXXXXKXGKXCDXXCNNAACXDXDC 31

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPREMBL 25:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_virus:*
- 16: sp_bacteriapi:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	98	70.5	638	13	042372 brachydanio
2	97	69.8	585	5	09U0E2
3	93	66.9	1194	13	09W737
4	91	65.5	2531	5	016004
5	90	64.7	191	5	08T520
6	90	64.7	191	5	08T521
7	90	64.7	193	5	08STG0
8	90	64.7	193	5	08T522
9	90	64.7	194	5	08T523
10	90	64.7	963	5	09GPM9
11	90	64.7	1270	5	09GPM0
12	90	64.7	1441	5	0867Q2
13	89	64.0	2528	13	08AXP0
14	88	63.3	1476	13	090285
15	87	62.6	2428	5	0816X6
16	85	61.2	892	5	025243

ALIGNMENTS

17	85	61.2	1455	5	086FJ9	086fj9 caenorhabdi
18	85	61.2	2516	11	07TQ52	07tq52 mus musculu
19	85	61.2	2526	11	07TQ51	07tq51 mus musculu
20	85	61.2	2531	11	08K428	08k428 mus musculu
21	85	61.2	2531	11	07TQ50	07tq50 mus musculu
22	85	61.2	2653	5	025253	025253 lucilia cup
23	83	59.7	762	13	042373	042373 brachydanio
24	83	59.7	2468	13	0800E4	0800e4 brachydanio
25	81	58.3	2524	5	09GPA5	09gpa5 branchiost
26	79	56.8	1308	5	09GPM8	09gpm8 caenorhabdi
27	78	56.1	847	4	09ULN3	09uln3 homo sapien
28	78	56.1	950	4	09ULN2	09uln2 homo sapien
29	78	56.1	3869	5	086PQ3	086pq3 cryptospori
30	77	55.4	260	13	09DEC9	09dec9 gallus gall
31	77	55.4	641	13	098U14	098u14 gallus gall
32	77	55.4	2352	5	061240	061240 halocynthia
33	76	54.7	752	13	042374	042374 brachydanio
34	74	53.2	40	5	09TXA3	09txa3 caenorhabdi
35	73	52.5	666	5	08SX14	08sx14 drosophila
36	73	52.5	666	5	09V553	09v553 drosophila
37	73	52.5	2447	13	013149	013149 fuqu rubrip
38	71	51.1	123	5	08T5X3	08t5x3 caenorhabdi
39	71	51.1	123	5	08T5X3	08t5x3 caenorhabdi
40	71	51.1	124	5	08T5X0	08t5x0 caenorhabdi
41	71	51.1	124	5	08T5W6	08t5w6 caenorhabdi
42	71	51.1	124	5	08T5W5	08t5w5 caenorhabdi
43	71	51.1	126	5	08T5W4	08t5w4 caenorhabdi
44	71	51.1	127	5	08T5W9	08t5w9 caenorhabdi
45	71	51.1	127	5	08T5X2	08t5x2 caenorhabdi

RESULT 1

042372 ID 042372 PRELIMINARY; PRT; 638 AA.

AC 042372; (TEMBREL. 05, Created)

DT 01-JUN-1998 (TEMBREL. 05, Last sequence update)

DT 01-JUN-1998 (TEMBREL. 25, Last annotation update)

DT 01-OCT-2003 (TEMBREL. 25, Last annotation update)

DE Notch receptor protein (Fragment).

GN NOTCH1B OR NOTCH1B.

OS Brachydanio rerio (Zebrafish) (Danio rerio).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio.

OC NCBI_TaxID=7955;

OX NCBI_TaxID=7955;

RN [1]

RP SEQUENCE FROM N.A.

RA Westin J., Lardelli M.;

RT "Three novel Notch genes in zebrafish: implications for vertebrate Notch gene evolution and function."

RL Dev. Genes Evol. 207:51-63(1997).

DR EMBL; Y10352; CAA71378.1; -.

DR HSSP; P00740; IEDM.

DR ZFIN; ZDB-GENE-990415-183; notch1b.

DR GO; GO:0016020; C:membrane; IEA.

DR GO; GO:0004872; F:receptor activity; IEA.

DR GO; GO:0030154; P:cell differentiation; IEA.

DR InterPro; IPR002110; ANK.

DR InterPro; IPR006209; EGF-like.

DR InterPro; IPR006210; IEGF.

DR InterPro; IPR008000; Notch_dom.

DR Pfam; PF00023; ank; 3.

DR Pfam; PF00068; EGF; 1.

DR Pfam; PF00666; notch; 3.

DR PRINTS; PRO1452; NOTCH.

DR SMART; SM00248; ANK; 2.

DR SMART; SM00181; EGF; 1.

DR SMART; SM00004; NL; 2.

DR PROSITE; PSS0088; ANK_REPEAT; 1.

DR PROSITE; PSS0297; ANK_REPEAT_REGION; 1.


```
DR PROSITE: PS00022; EGF_1; 2.
KM ANK repeat; EGF-like domain; Receptor; Repeat.
FT NON_TER 1
SQ SEQUENCE 638 AA; 70800 MW; BCBAL6D7561FA3D2 CRC64;

Query Match 70.5%; Score 98; DB 13; Length 638;
Best Local Similarity 51.6%; Pred. No. 1.1e-08;
Matches 16; Conservative 1; Mismatches 14; Indels 0; Gaps 0;

QY 1 CDXXXXXXGNGXCDXXCNNAACXXDXGDC 31
DB 86 CEIEQCKVKGKNCIKDSACNNYACDWDGDC 116

RESULT 2
Q9U0E2 PRELIMINARY; PRT; 585 AA.
ID Q9U0E2
DT 01-MAY-2000 (TRENBLREL. 13, Created)
DT 01-MAY-2000 (TRENBLREL. 13, Last sequence update)
DT 01-OCT-2003 (TRENBLREL. 25, Last annotation update)
DE Signal receptor protein (Fragment).
GN NOTCH.
OS Tribolium castaneum (Red flour beetle).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Coleoptera; Polyphaga; Cucujiformia;
OC Tenebrionidae; Tribolium.
OX NCBI_TaxID=7070;
RN [1]
RP Tautz D., Lardelli M., Westin J., Tamme R.;
RT "Embryonic expression of Tribolium Notch.";
RT Submitted (Apr-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ005083; CAB65469.1; -.
DR HSSP; P00740; 1EDM.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005509; F:calcium ion binding; IEA.
DR GO; GO:0004872; F:receptor activity; IEA.
DR GO; GO:0030154; P:cell differentiation; IEA.
DR InterPro; IPR000152; Asx hydroxyl_S.
DR InterPro; IPR000742; EGF_2.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR001438; EGF_11.
DR InterPro; IPR006209; EGF-like.
DR InterPro; IPR000800; Notch_dom.
DR Pfam; PF00008; EGF_12.
DR Pfam; PF00066; notch; 2.
DR PRINTS; PR00010; EGFLOOD.
DR PRINTS; PR01452; NOTCH.
DR SMART; SM00179; EGF_CA; 5.
DR SMART; SM00004; NL; 2.
DR PROSITE; PS00010; ASX HYDROXYL; 6.
DR PROSITE; PS00022; EGF_1; 12.
DR PROSITE; PS0186; EGF_2; 9.
DR PROSITE; PS0187; EGF_CA; 4.
KM EGF-like domain; Receptor.
FT NON_TER 1
FT NON_TER 585
SQ SEQUENCE 585 AA; 63337 MW; 4CF7A51D0820D048 CRC64;

Query Match 69.8%; Score 97; DB 5; Length 585;
Best Local Similarity 51.6%; Pred. No. 1.6e-08;
Matches 16; Conservative 1; Mismatches 14; Indels 0; Gaps 0;

QY 1 CDXXXXXXGNGXCDXXCNNAACXXDXGDC 31
DB 488 CNGKCPYKRGNGRCDSECNYYACGFDGDC 518

RESULT 3
Q9W737 PRELIMINARY; PRT; 1194 AA.
ID Q9W737
```

```
AC Q9W737;
DT 01-NOV-1999 (TRENBLREL. 12, Created)
DT 01-NOV-1999 (TRENBLREL. 12, Last sequence update)
DT 01-OCT-2003 (TRENBLREL. 25, Last annotation update)
DE Notch-1 (Fragment).
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=99328644; PubMed=10402194;
RA Wakamatsu Y., Maynard T.M., Jones S.U., Weston J.A.;
RT "NTMB localizes in the basal cortex of mitotic avian neuroepithelial
RT cells and modulates neuronal differentiation by binding to NOTCH-1.";
RL Neuron 23:71-81 (1999).
DR EMBL; AF159231; AAD42893.1; -.
DR HSSP; P00740; 1EDM.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0030154; P:cell differentiation; IEA.
DR InterPro; IPR002110; ANK.
DR InterPro; IPR001064; Crystallin.
DR InterPro; IPR006209; EGF-like.
DR InterPro; IPR006210; IEGF.
DR InterPro; IPR000800; Notch_dom.
DR Pfam; PF00023; ank; 6.
DR Pfam; PF00008; EGF; 1.
DR Pfam; PF00066; notch; 3.
DR PRINTS; PR01452; NOTCH.
DR SMART; SM00248; ANK; 5.
DR SMART; SM00181; EGF; 1.
DR SMART; SM00004; NL; 3.
DR PROSITE; PS00086; ANK_REPEAT; 4.
DR PROSITE; PS0297; ANK_REPEAT_REGION; 1.
DR PROSITE; PS00225; CRYSTALLIN_BETAGAMMA; 1.
DR PROSITE; PS00022; EGF_1; 2.
KM ANK repeat; Repeat.
FT NON_TER 1
SQ SEQUENCE 1194 AA; 130641 MW; 650380B86584974 CRC64;

Query Match 66.9%; Score 93; DB 13; Length 1194;
Best Local Similarity 48.4%; Pred. No. 1.4e-07;
Matches 15; Conservative 1; Mismatches 15; Indels 0; Gaps 0;

QY 1 CDXXXXXXGNGXCDXXCNNAACXXDXGDC 31
DB 92 CEIACASVAGNKICDCKCNHACGWDGDC 122

RESULT 4
O16004 PRELIMINARY; PRT; 2531 AA.
ID O16004
AC O16004;
DT 01-JAN-1998 (TRENBLREL. 05, Created)
DT 01-JAN-1998 (TRENBLREL. 05, Last sequence update)
DT 01-OCT-2003 (TRENBLREL. 25, Last annotation update)
DE Notch homolog.
OS Lytechinus variegatus (Sea urchin).
OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC Echinoidea; Euechinoidea; Echinacea; Temnopneuroidea; Toxopneustidae;
OC Lytechinidae.
OX NCBI_TaxID=7654;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97454256; PubMed=9310331;
RA Sherwood D.R., McClay D.R.;
RT "Identification and localization of a sea urchin Notch homologue;
RT insights into vegetal plate regionalization and Notch receptor
RT regulation.";
RL Development 124:3363-3374 (1997).
DR EMBL; AF000634; AAB82088.1; -.
```

DR HSP; P01132; 1EGF.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005509; F:calcium ion binding; IEA.
DR GO; GO:0030154; P:cell differentiation; IEA.
DR InterPro; IPR002110; ANK.
DR InterPro; IPR000152; Asx_hydroxyl_S.
DR InterPro; IPR000742; EGF_2.
DR InterPro; IPR001881; EGF_2.
DR InterPro; IPR001438; EGF_11.
DR InterPro; IPR006209; EGF_11.
DR InterPro; IPR008297; Notch.
DR InterPro; IPR008000; Notch_dom.
DR Pfam; PF00023; ank; 6.
DR Pfam; PF00008; EGF; 34.
DR Pfam; PF00066; notch; 3.
DR PRINTS; PR00110; EGFALOOD.
DR PRINTS; PR01452; NOTCH.
DR SMART; SM00248; ANK; 5.
DR SMART; SM00179; EGF_CA; 23.
DR SMART; SM00004; NL; 3.
DR PROSITE; PS50088; ANK_REPEAT; 5.
DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.
DR PROSITE; PS00010; ASX_HYDROXYL; 21.
DR PROSITE; PS00022; EGF_1; 33.
DR PROSITE; PS01186; EGF_2; 25.
DR PROSITE; PS01187; EGF_CA; 20.
DR PIRSF; PIRSF002279; Notch; 1.
DR ANK repeat; EGF-like domain; Repeat.
SQ SEQUENCE 2531 AA; 273982 MW; 5BFA2BEC627CA303 CRC64;

Query Match 65.5%; Score 91; DB 5; Length 2531;
Best Local Similarity 53.8%; Pred. No. 6.3e-07;
Matches 14; Conservative 1; Mismatches 11; Indels 0; Gaps 0;

Oy 6 CXXXXXGNGXCDXXCNNAACXXDXDC 31
Db 1526 CLERKNGFCDEBCNNTGLYDGLDC 1551

RESULT 5
O8T5Z0 PRELIMINARY; PRT; 191 AA.
AC O8T5Z0.
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Glp-1 (Fragment).
OS Caenorhabditis briggsae.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6238;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=V7847;
RA Grunstein A., Gaspar J.M., Walters J.R., Palopoli M.F.,
RT "Levels of DNA polymorphism vary with mating system in the nematode
genus Caenorhabditis."
RL Genetics 0:0-0(2002).
DR EMBL; AF491463; AAM09703.1; -
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR GO; GO:0030154; P:cell differentiation; IEA.
DR InterPro; IPR000152; Asx_hydroxyl_S.
DR InterPro; IPR006209; EGF_like.
DR InterPro; IPR006210; IEGF.
DR InterPro; IPR002049; Laminin_EGF.
DR InterPro; IPR000800; Notch_dom.
DR Pfam; PF00008; EGF; 3.
DR Pfam; PF00066; notch; 2.
DR PRINTS; PR00011; EGFAMININ.
DR PRINTS; PR01452; NOTCH.
DR SMART; SM00181; EGF; 3.
DR SMART; SM00004; NL; 2.

DR PROSITE; PS00010; ASX_HYDROXYL; 1.
DR PROSITE; PS00022; EGF_1; 3.
DR PROSITE; PS01186; EGF_2; 3.
KW EGF-like domain.
FT NON TER 1
FT NON TER 1
SQ SEQUENCE 191 AA; 20943 MW; 2EBB21B5FA47FB470 CRC64;

Query Match 64.7%; Score 90; DB 5; Length 191;
Best Local Similarity 48.4%; Pred. No. 9e-08;
Matches 15; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

Oy 1 CDXXXXXGNGXCDXXCNNAACXXDXDC 31
Db 117 CKKNCALAGNGICDEDCNYAEQFDGDC 147

RESULT 6
O8T5Z1 PRELIMINARY; PRT; 191 AA.
AC O8T5Z1.
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Glp-1 (Fragment).
OS Caenorhabditis briggsae.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6238;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PB800;
RA Grunstein A., Gaspar J.M., Walters J.R., Palopoli M.F.,
RT "Levels of DNA polymorphism vary with mating system in the nematode
genus Caenorhabditis."
RL Genetics 0:0-0(2002).
DR EMBL; AF491462; AAM09702.1; -
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR GO; GO:0030154; P:cell differentiation; IEA.
DR InterPro; IPR000152; Asx_hydroxyl_S.
DR InterPro; IPR006209; EGF_like.
DR InterPro; IPR006210; IEGF.
DR InterPro; IPR002049; Laminin_EGF.
DR InterPro; IPR000800; Notch_dom.
DR Pfam; PF00008; EGF; 3.
DR Pfam; PF00066; notch; 2.
DR PRINTS; PR00011; EGFAMININ.
DR PRINTS; PR01452; NOTCH.
DR SMART; SM00181; EGF; 3.
DR SMART; SM00004; NL; 2.
DR PROSITE; PS00010; ASX_HYDROXYL; 1.
DR PROSITE; PS00022; EGF_1; 3.
DR PROSITE; PS01186; EGF_2; 3.
KW EGF-like domain.
FT NON TER 1
FT NON TER 1
SQ SEQUENCE 191 AA; 20977 MW; 2EBB21B5FFA46470 CRC64;

Query Match 64.7%; Score 90; DB 5; Length 191;
Best Local Similarity 48.4%; Pred. No. 9e-08;
Matches 15; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

Oy 1 CDXXXXXGNGXCDXXCNNAACXXDXDC 31
Db 117 CKKNCALAGNGICDEDCNYAEQFDGDC 147

RESULT 7
O8T5Z0 PRELIMINARY; PRT; 193 AA.
AC O8T5Z0.
DT 01-JUN-2002 (TrEMBLrel. 21, Created)

DT 01-JUN-2002 (TREMblrel. 21, last sequence update)
DT 01-OCT-2003 (TREMblrel. 25, last annotation update)
DE Glp-1 (Fragment).
OS Caenorhabditis briggsae.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6238;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HK104, and HK105;
RA Graustein A., Gaspar J.M., Walters J.R., Palopoli M.F.;
RT "Levels of DNA polymorphism vary with mating system in the nematode
genus Caenorhabditis."
RT Genetics 0:0-0(2002).
DR EMBL; AF491459; AAM0699.1; -.
DR EMBL; AF491461; AAM09701.1; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR GO; GO:0030154; P:cell differentiation; IEA.
DR InterPro; IPR000152; Asx hydroxyl_S.
DR InterPro; IPR006209; EGF-like.
DR InterPro; IPR002049; Laminin_EGF.
DR Pfam; PF00008; EGF; 3.
DR Pfam; PF00066; notch; 2.
DR PRINTS; PR00011; EGF_LAMININ.
DR PRINTS; PR01452; NOTCH.
DR SMART; SM00181; EGF; 3.
DR SMART; SM00004; NL; 2.
DR PROSITE; PS00010; ASX_HYDROXYL; 1.
DR PROSITE; PS00022; EGF_1; 3.
DR PROSITE; PS01186; EGF_2; 3.
KW EGF-like domain.
FT NON_TER 1 1
FT NON_TER 193 193
SQ SEQUENCE 193 AA, 21180 MW, 41308AB647B5C540 CRC64;

Query Match 64.7%; Score 90; DB 5; Length 193;
Best Local Similarity 48.4%; Pred. No. 9.1e-08;
Matches 15; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY 1 CDXXXCKXKXGKCDXXCNNAACXXDXDC 31
DB 120 CKKXCKALAGNIGCDPCNYACCPDGGDC 150

RESULT 8
ID Q8T5Z2 PRELIMINARY; PRT; 193 AA.
AC Q8T5Z2;
DT 01-JUN-2002 (TREMblrel. 21, Created)
DT 01-JUN-2002 (TREMblrel. 21, last sequence update)
DT 01-OCT-2003 (TREMblrel. 25, last annotation update)
DE Glp-1 (Fragment).
OS Caenorhabditis briggsae.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6238;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AF16;
RA Graustein A., Gaspar J.M., Walters J.R., Palopoli M.F.;
RT "Levels of DNA polymorphism vary with mating system in the nematode
genus Caenorhabditis."
RT Genetics 0:0-0(2002).
DR EMBL; AF491460; AAM09700.1; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR GO; GO:0030154; P:cell differentiation; IEA.
DR InterPro; IPR000152; Asx hydroxyl_S.
DR InterPro; IPR006209; EGF-like.
DR InterPro; IPR006210; IEGF.

DR InterPro; IPR002049; Laminin_EGF.
DR InterPro; IPR000800; Notch_dom.
DR Pfam; PF00008; EGF; 3.
DR Pfam; PF00066; notch; 2.
DR PRINTS; PR00011; EGF_LAMININ.
DR PRINTS; PR01452; NOTCH.
DR SMART; SM00181; EGF; 3.
DR SMART; SM00004; NL; 2.
DR PROSITE; PS00010; ASX_HYDROXYL; 1.
DR PROSITE; PS00022; EGF_1; 3.
DR PROSITE; PS01186; EGF_2; 3.
KW EGF-like domain.
FT NON_TER 1 1
FT NON_TER 193 193
SQ SEQUENCE 193 AA, 21178 MW, 42135B88B8E02C5 CRC64;

Query Match 64.7%; Score 90; DB 5; Length 193;
Best Local Similarity 48.4%; Pred. No. 9.1e-08;
Matches 15; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY 1 CDXXXCKXKXGKCDXXCNNAACXXDXDC 31
DB 119 CKKXCKALAGNIGCDPCNYACCPDGGDC 149

RESULT 9
ID Q8T5Z3 PRELIMINARY; PRT; 194 AA.
AC Q8T5Z3;
DT 01-JUN-2002 (TREMblrel. 21, Created)
DT 01-JUN-2002 (TREMblrel. 21, last sequence update)
DT 01-OCT-2003 (TREMblrel. 25, last annotation update)
DE Glp-1 (Fragment).
OS Caenorhabditis briggsae.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6238;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PB826;
RA Graustein A., Gaspar J.M., Walters J.R., Palopoli M.F.;
RT "Levels of DNA polymorphism vary with mating system in the nematode
genus Caenorhabditis."
RT Genetics 0:0-0(2002).
DR EMBL; AF491458; AAM06998.1; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR GO; GO:0030154; P:cell differentiation; IEA.
DR InterPro; IPR000152; Asx hydroxyl_S.
DR InterPro; IPR006209; EGF-like.
DR InterPro; IPR006210; IEGF.
DR InterPro; IPR002049; Laminin_EGF.
DR InterPro; IPR000800; Notch_dom.
DR Pfam; PF00008; EGF; 3.
DR Pfam; PF00066; notch; 2.
DR PRINTS; PR00011; EGF_LAMININ.
DR PRINTS; PR01452; NOTCH.
DR SMART; SM00181; EGF; 3.
DR SMART; SM00004; NL; 2.
DR PROSITE; PS00010; ASX_HYDROXYL; 1.
DR PROSITE; PS00022; EGF_1; 3.
DR PROSITE; PS01186; EGF_2; 3.
KW EGF-like domain.
FT NON_TER 1 1
FT NON_TER 194 194
SQ SEQUENCE 194 AA, 21275 MW, 07313C8ABD9C65C5 CRC64;

Query Match 64.7%; Score 90; DB 5; Length 194;
Best Local Similarity 48.4%; Pred. No. 9.1e-08;
Matches 15; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY 1 CDXXXCKXKXGKCDXXCNNAACXXDXDC 31

DB 120 CCKNCKALAGNGICDEDCNVAEQFDGDC 150

RESULT 10

Q9GPM9 PRELIMINARY; PRT; 963 AA.

ID Q9GPM9

AC 01-MAR-2001 (Tremblrel. 16, Created)

DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)

DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)

DE Notch-like transmembrane receptor (Fragment).

GN GLP-1.

OS Caenorhabditis briggsae.

OC Eukaryota; Metazoa; Nematoda; Chromodorea; Rhabditida; Rhabditoidea;

OC Rhabditidae; Peloderinae; Caenorhabditis.

OX NCBI_Taxid=6238;

SEQUENCE FROM N.A.

RP STRAIN=AF16;

RX MEDLINE=21100341; PubMed=11156985;

RA "Conservation of glp-1 regulation and function in nematodes.";

RL Genetics 157:639-654 (2001).

DR EMBL; AF315555; AAG49317.1; -.

DR HSSP; P01132; 1EGF.

DR GO; GO:0016021; C:integral to membrane; IEA.

DR GO; GO:0004872; F:receptor activity; IEA.

DR GO; GO:0005198; F:structural molecule activity; IEA.

DR GO; GO:0030154; P:cell differentiation; IEA.

DR InterPro; IPR002110; ANK.

DR InterPro; IPR000152; Axx hydroxyl_S.

DR InterPro; IPR000742; EGF_2.

DR InterPro; IPR006209; EGF-like.

DR InterPro; IPR006210; 1EGF.

DR InterPro; IPR002049; Laminin EGF.

DR InterPro; IPR000800; Notch_dom.

DR Pfam; PF00023; ank; 6.

DR Pfam; PF00066; EGF; 4.

DR Pfam; PF00066; notch; 3.

DR PRINTS; PR00011; EGF/LAMININ.

DR PRINTS; PR01452; NOTCH.

DR SMART; SM00248; ANK; 4.

DR SMART; SM00181; EGF; 4.

DR SMART; SM00004; NL; 3.

DR PROSITE; PS50088; ANK_REPEAT; 3.

DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.

DR PROSITE; PS00010; ASX_HYDROXYL; 1.

DR PROSITE; PS00022; EGF_1; 4.

DR PROSITE; PS01186; EGF_2; 3.

KW ANK repeat; EGF-like domain; Receptor; Repeat; Transmembrane.

FT NON TER

SO SEQUENCE 963 AA; 106394 MW; 52BF80010E12F857 CRC64;

Query Match 64.7%; Score 90; DB 5; Length 963;

Best Local Similarity 48.4%; Pred. No. 3.9e-07;

Matches 15; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

Qy 1 CDXXXCKKNGKNGKCDXXCNNAACXXDXDC 31

Db 187 CCKNCKALAGNGICDEDCNVAEQFDGDC 217

RESULT 11

Q9GPM9 PRELIMINARY; PRT; 1270 AA.

ID Q9GPM9

AC 01-MAR-2001 (Tremblrel. 16, Created)

DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)

DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)

DE Notch-like transmembrane receptor.

GN GLP-1.

OS Caenorhabditis briggsae.

OC Eukaryota; Metazoa; Nematoda; Chromodorea; Rhabditida; Rhabditoidea;

OC Rhabditidae; Peloderinae; Caenorhabditis.

OX NCBI_Taxid=6238;

RP SEQUENCE FROM N.A.

RC STRAIN=AF16;

RX MEDLINE=21100341; PubMed=11156985;

RA "Conservation of glp-1 regulation and function in nematodes.";

RL Genetics 157:639-654 (2001).

DR EMBL; AF315554; AAG49316.1; -.

DR HSSP; P01132; 1EGF.

DR GO; GO:0016021; C:integral to membrane; IEA.

DR GO; GO:0005509; F:calcium ion binding; IEA.

DR GO; GO:0004872; F:receptor activity; IEA.

DR GO; GO:0005198; F:structural molecule activity; IEA.

DR GO; GO:0030154; P:cell differentiation; IEA.

DR InterPro; IPR002110; ANK.

DR InterPro; IPR000152; Axx hydroxyl_S.

DR InterPro; IPR000742; EGF_2.

DR InterPro; IPR001881; EGF_CA.

DR InterPro; IPR006209; EGF-like.

DR InterPro; IPR002049; Laminin EGF.

DR InterPro; IPR000800; Notch_dom.

DR Pfam; PF00023; ank; 6.

DR Pfam; PF00008; EGF; 10.

DR Pfam; PF00066; notch; 3.

DR PRINTS; PR00011; EGF/LAMININ.

DR PRINTS; PR01452; NOTCH.

DR SMART; SM00248; ANK; 4.

DR SMART; SM00179; EGF_CA; 1.

DR SMART; SM00004; NL; 3.

DR PROSITE; PS50088; ANK_REPEAT; 3.

DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.

DR PROSITE; PS00010; ASX_HYDROXYL; 3.

DR PROSITE; PS00022; EGF_1; 10.

DR PROSITE; PS01186; EGF_2; 8.

DR PROSITE; PS01187; EGF_CA; 1.

KW ANK repeat; EGF-like domain; Receptor; Repeat; Transmembrane.

SO SEQUENCE 1270 AA; 138964 MW; A7662EB575A4B61B CRC64;

Query Match 64.7%; Score 90; DB 5; Length 1270;

Best Local Similarity 48.4%; Pred. No. 5e-07;

Matches 15; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

Qy 1 CDXXXCKKNGKNGKCDXXCNNAACXXDXDC 31

Db 494 CCKNCKALAGNGICDEDCNVAEQFDGDC 524

RESULT 12

Q867Q2 PRELIMINARY; PRT; 1441 AA.

ID Q867Q2

AC 0867Q2;

DT 01-JUN-2003 (Tremblrel. 24, Created)

DT 01-JUN-2003 (Tremblrel. 24, Last sequence update)

DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)

DE Notch-like transmembrane receptor LIN-12.

OS Caenorhabditis briggsae.

OC Eukaryota; Metazoa; Nematoda; Chromodorea; Rhabditida; Rhabditoidea;

OC Rhabditidae; Peloderinae; Caenorhabditis.

OX NCBI_Taxid=6238;

SEQUENCE FROM N.A.

RP MEDLINE=22243075; PubMed=12356262;

RX "Evolution of discrete notch-like receptors from a distant gene duplication in Caenorhabditis.";

RT EMBL; AF499438; AAP05765.1; -.

RT EMBL; AF499440; AAP05765.1; -.

DR GO; GO:0016021; C:integral to membrane; IEA.

DR GO; GO:0005509; F:calcium ion binding; IEA.

DR GO; GO:0004872; F:receptor activity; IEA.

DR TTSIB=Retina;
 RA MEDLINE=9735970; PubMed=9216061;
 RA Sullivan S.A., Barchel L.K., Largent B.L., Raymond P.A.,
 RT "A goldfish Notch-3 homologue is expressed in neurogenic regions of
 RT embryonic, adult, and regenerating brain and retina";
 RL Dev. Genet. 20:208-223(1997).
 DR EMBL: U0191; AAB17010.2; -.
 DR HSSP: P01137; 1EGF.
 DR GO: GO:0016020; Cmembrane, IEA.
 DR GO: GO:0005509; Fcalcium ion binding, IEA.
 DR GO: GO:0030154; Pcell differentiation, IEA.
 DR InterPro: IPR002110; ANK.
 DR InterPro: IPR000152; Asx hydroxyl_S.
 DR InterPro: IPR000742; EGF 2.
 DR InterPro: IPR001881; EGF Ca.
 DR InterPro: IPR001438; EGF II.
 DR InterPro: IPR006289; EGF-like.
 DR InterPro: IPR000800; Notch_dom.
 DR Pfam: PF00023; ank, 6.
 DR Pfam: PF00068; EGF, 11.
 DR Pfam: PF00066; notch, 3.
 DR PRINTS: PR00010; EGFBLD.
 DR PRINTS: PR01452; NOTCH.
 DR SMART: SM00248; ANK; 5.
 DR SMART: SM00179; EGF CA; 4.
 DR SMART: SM00004; NT; 3.
 DR PROSITE: PS50088; ANK_REPEAT, 4.
 DR PROSITE: PS50297; ANK_REPEAT_REGION, 1.
 DR PROSITE: PS00010; ASX_HYDROXYL, 3.
 DR PROSITE: PS00022; EGF 1, 10.
 DR PROSITE: PS01186; EGF 2; 9.
 DR PROSITE: PS01187; EGF CA; 2.

KM ANK repeat; EGF-like domain; Repeat.
 FT NON TER 1
 FT NON TER 1476 1476
 SQ SEQUENCE 1476 AA; 160385 MW; 03352B906245A283 CRC64;

Query Match 63.3%; Score 88; DB 13; Length 1476;
 Best Local Similarity 45.2%; Pred. No. 1.3e-06;
 Matches 14; Conservative 2; Mismatches 15; Indels 0; Gaps 0;

OY 1 CDXXXXXGNGXCDXXCNNAACXXDGXDC 31
 Db 449 CP1AECHEKANDSVCDKCNLSLACRWDGDC 479

RESULT 15
 ID 0816X6 PRELIMINARY; PRT; 2428 AA.
 AC 0816X6;
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Notch-like protein.
 GN BOTCH.
 OS Boophilus microplus (Cattle tick).
 OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;
 OC Parasitiformes; Ixodida; Ixodidae; Boophilus.
 OX NCBI_TaxID=6941;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Whole ticks;
 RA Whitefield P.L., McNicholas P.;
 RT "BOTCH, A Notch Homolog from the Cattle Tick Boophilus microplus.";
 RL Submitted (Aug-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF537369; AAN06819.1; -
 DR GO; GO:0016020; C:membrane; IEA.
 DR GO; GO:0005509; F:calcium ion binding; IEA.
 DR GO; GO:0005198; F:structural molecule activity; IEA.
 DR GO; GO:0030154; P:cell differentiation; IEA.
 DR InterPro: IPR002110; ANK.
 DR InterPro: IPR000152; Asx_hydroxyl_S.
 DR InterPro: IPR000742; EGF_2.
 DR InterPro: IPR001881; EGF_Ca.
 DR InterPro: IPR001438; EGF_II.
 DR InterPro: IPR006209; EGF_like.
 DR InterPro: IPR006210; IEGF.
 DR InterPro: IPR002048; laminin_EGF.
 DR InterPro: IPR008297; Notch.
 DR InterPro: IPR008000; Notch_dom.
 DR Pfam; PF00023; ank; 6.
 DR Pfam; PF00008; EGF; 35.
 DR Pfam; PF00066; notch; 3.
 DR PRINTS; PR00010; EGFBL00D.
 DR PRINTS; PR00011; EGFPLAMININ.
 DR PRINTS; PR01452; NOTCH.
 DR SMART; SM00248; ANK; 7.
 DR SMART; SM00181; EGF; 36.
 DR SMART; SM00179; EGF_Ca; 33.
 DR SMART; SM00004; NL; 3.
 DR PROSITE; PS50088; ANK_REPEAT; 4.
 DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.
 DR PROSITE; PS00010; ASX_HYDROXYL; 22.
 DR PROSITE; PS00022; EGF_1; 34.
 DR PROSITE; PS01186; EGF_2; 27.
 DR PROSITE; PS01187; EGF_Ca; 22.
 DR PIRSF; PIRSF002279; Notch; 1.
 SQ SEQUENCE 2428 AA; 260842 MW; 766A9362CE37CB9F CRC64;

Query Match 62.6%; Score 87; DB 5; Length 2428;
 Best Local Similarity 53.8%; Pred. No. 3e-06;
 Matches 14; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

OY 6 CXXKXGNGXCDXXCNNAACXXDGXDC 31

DB 1541 CMRYGNGYCDQGCNNECNWIDGDC 1566
 Search completed: May 19, 2004, 15:59:51
 Job time : 3.98606 secs